Supplemental Material for "Weak-Identification Robust Wild Bootstrap applied to a Consistent Model Specification Test"

Jonathan B. Hill*
University of North Carolina – Chapel Hill

March 25, 2020

Contents

*Dept. of Economics, University of North Carolina, Chapel Hill; www.unc.edu/~jbhill; jbhill@email.unc.edu

List of Tables

A Outline and Assumptions

Appendix B contains proofs of the supporting lemmata from the main paper. In Appendix C we prove Theorem 4.1. Appendix D details the Identification Category Selection Type 2 [ICS-2] p-value. Appendix E presents bootstrapped identification category robust critical values, with asymptotic theory. Assumptions 3-5 are discussed in Appendix F in the context of a STAR model.

Recall the model

$$y_t = \zeta_0' x_t + \beta_0' g(x_t, \pi_0) + \epsilon_t = f(\theta_0, x_t) + \epsilon_t \text{ where } x_t \in \mathbb{R}^{k_x} \text{ and } \theta \equiv [\zeta', \beta', \pi']' . \tag{A.1}$$

The variable y_t is a scalar, $x_t \in \mathbb{R}^{k_x}$ are covariates with finite $k_x \geq 2$, $g : \mathbb{R}^{k_x} \times \Pi \to \mathbb{R}^{k_\beta}$ is a known function, and $\zeta_0 \in \mathcal{Z}$, $\beta_0 \in \mathcal{B}$ and $\pi_0 \in \Pi$, where $\mathcal{B}$, $\mathcal{Z}$ and Π are compact subsets of $\mathbb{R}^{k_\beta}$, $\mathbb{R}^{k_x}$ and $\mathbb{R}^{k_\pi}$ respectively for finite $k_\pi \geq 1$. The covariates x_t include a constant term and at least one stochastic regressor. Assume $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] \in (0, \infty)$ for some unique $\theta_0 \in \Theta \equiv \mathcal{Z} \times \mathcal{B} \times \Pi$.

Let y_t exist on the probability measure space $(\Omega, \mathcal{P}, \mathcal{F})$, where $\mathcal{F} \equiv \sigma(\cup_{t \in \mathbb{Z}} \mathcal{F}_t)$ and $\mathcal{F}_t \equiv \sigma(y_\tau : \tau \leq t)$. Assume Θ has the form $\{\theta \equiv [\beta', \zeta', \pi']' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\}$, where $\mathcal{B}$, $\mathcal{Z}(\beta)$ for each β, and Π are compact subsets. Recall:

$$\psi \equiv [\beta', \zeta']' \in \Psi \equiv \{(\beta, \zeta) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta)\}.$$

The true parameter space $\Theta^* = \Psi^* \times \Pi^* = \{\theta \equiv [\beta', \zeta', \pi']' : \beta \in \mathcal{B}^*, \zeta \in \mathcal{Z}^*(\beta), \pi \in \Pi^*\}$ lies in the interior of Θ, it contains $\theta_0 \equiv [\beta_0', \zeta_0', \pi_0']'$, and $0 \in \mathcal{B}^*$.

Recall the following definitions and constructions:

$$\mathfrak{B}(\beta) = \begin{bmatrix} I_{k_\psi} & 0_{k_\psi \times 2} \\ 0_{2 \times k_\psi} & \|\beta\| \times I_2 \end{bmatrix}, \tag{A.2}$$

and

$$\omega(\beta) \equiv \begin{cases} \beta / \|\beta\| & \text{if } \beta \neq 0 \\ 1_{k_\beta} / \left\| 1_{k_\beta} \right\| & \text{if } \beta = 0 \end{cases},$$

and

$$\begin{aligned} d_{\psi,t}(\pi) &\equiv [g(x_t, \pi)', x_t']' \\ d_{\theta,t}(\omega, \pi) &\equiv \left[g(x_t, \pi)', x_t', \omega' \frac{\partial}{\partial \pi} g(x_t, \pi) \right]' \\ d_{\theta,t} &\equiv d_{\theta,t}(\omega_0, \pi_0) \end{aligned}$$

$$
\begin{aligned}
&\mathfrak{b}_{\psi}(\pi,\lambda) = E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\psi,t}(\pi)\right]\\
&\mathfrak{b}_{\theta}(\omega,\pi,\lambda) \equiv E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\theta,t}(\omega,\pi)\right]\\
&\mathfrak{b}_{\theta}(\lambda) \equiv E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\theta,t}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{H}_{\psi}(\pi) \equiv E\left[d_{\psi,t}(\pi)d_{\psi,t}(\pi)'\right]\\
&\mathcal{H}_{\theta}(\omega,\pi) \equiv E\left[d_{\theta,t}(\omega,\pi)d'_{\theta,t}(\omega,\pi)\right]\\
&\mathcal{H}_{\theta} \equiv \mathcal{H}_{\theta}(\omega_0,\pi_0) = E\left[d_{\theta,t}d'_{\theta,t}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{K}_{\psi,t}(\pi,\lambda) \equiv F\left(\lambda'W(x_t)\right) - \mathfrak{b}_{\psi}(\pi,\lambda)'\mathcal{H}_{\psi}^{-1}(\pi)d_{\psi,t}(\pi)\\
&\mathcal{K}_{\theta,t}(\lambda) \equiv F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_{\theta}(\lambda)'\mathcal{H}_{\theta}^{-1}d_{\theta,t}(\beta_n/\left\|\beta_n\right\|,\pi_0)\\
&\mathcal{K}_{\theta,t}(\lambda;a,m) \equiv \sum\nolimits_{i=1}^{m}\alpha_i\mathcal{K}_{\theta,t}(\lambda_i),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{G}_{\psi,n}(\theta) &= \sqrt{n}\left\{\frac{\partial}{\partial\psi}Q_n(\theta) - E\left[\frac{\partial}{\partial\psi}Q_n(\theta)\right]\right\} \qquad (A.3)\\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\theta)d_{\psi,t}(\pi) - E\left[\epsilon_t(\theta)d_{\psi,t}(\pi)\right]\right\}\\
\mathcal{G}_{\theta,n}(\theta) &= \mathfrak{B}(\beta_n)^{-1}\sqrt{n}\left\{\frac{\partial}{\partial\theta}Q_n(\theta) - E\left[\frac{\partial}{\partial\theta}Q_n(\theta)\right]\right\}\\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\theta)d_{\theta,t}(\omega(\beta),\pi) - E\left[\epsilon_t(\theta)d_{\theta,t}(\omega(\beta),\pi)\right]\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathcal{D}_{\psi}(\pi) \equiv -\frac{\partial}{\partial\beta_0'}E\left[\epsilon_t(\theta)d_{\psi,t}(\pi)\right] = -E\left[d_{\psi,t}(\pi)g(x_t,\pi_0)'\right] \qquad (A.4)\\
&\mathcal{H}_{\psi}(\pi) \equiv E\left[d_{\psi,t}(\pi)d_{\psi,t}(\pi)'\right],
\end{aligned}
$$

and

$$
\begin{aligned}
&\widehat{\mathcal{H}}_n = \frac{1}{n}\sum_{t=1}^{n}d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)' \text{ where } \omega(\beta) \equiv \begin{cases} \beta/\left\|\beta\right\| & \text{if } \beta \neq 0 \\ 1_{k_\beta}/\left\|1_{k_\beta}\right\| & \text{if } \beta = 0 \end{cases} \qquad (A.5)\\
&\widehat{\mathcal{H}}_{\psi,n}(\pi) \equiv \frac{1}{n}\sum_{t=1}^{n}d_{\psi,t}(\pi)d_{\psi,t}(\pi)'
\end{aligned}
$$

$$
\begin{aligned}
&\hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n} F\left(\lambda'\mathcal{W}(x_t)\right) d_{\theta,t}(\omega,\pi) \\
&\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n} F\left(\lambda'\mathcal{W}(x_t)\right) d_{\psi,t}(\pi) \;\text{ and }\; \mathfrak{b}_{\psi}(\pi,\lambda) \equiv E\left[F\left(\lambda'\mathcal{W}(x_t)\right) d_{\psi,t}(\pi)\right] \\
&\hat{v}_n^2(\hat{\theta}_n,\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\hat{\theta}_n)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\theta,n}(\omega(\hat{\beta}_n),\hat{\pi}_n,\lambda)'\widehat{\mathcal{H}}_n^{-1} d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)\right\}^2 \\
&\hat{\mathcal{V}}_n \equiv \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\hat{\theta}_n) d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n) d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n) \\
&\hat{\Sigma}_n \equiv \widehat{\mathcal{H}}_n^{-1}\hat{\mathcal{V}}_n\widehat{\mathcal{H}}_n^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}_{\psi,n}(\pi;a,r) \equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\sum_{i=1}^{m}\alpha_i r' d_{\psi,t}(\pi_i) \\
&\mathbb{E}_{\theta,n}(\omega,\pi;a,r) \equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\sum_{i=1}^{m}\alpha_i r' d_{\theta,t}(\omega_i,\pi_i) \\
&\mathfrak{EG}_{\psi,n}(\lambda;a,r) \equiv r_1\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\alpha_i\left\{\epsilon_t(\psi_n,\pi_i)\mathcal{K}_{\psi,t}(\pi_i,\lambda_i) - E\left[\epsilon_t(\psi_n,\pi_i)\mathcal{K}_{\psi,t}(\pi_i,\lambda_i)\right]\right\} \\
&\qquad\qquad + r_2'\sum_{i=1}^{m}\alpha_i\mathcal{G}_{\psi,n}(\psi_n,\pi_i).
\end{aligned}
$$

Recall the statistic used to determine whether b is finite:

$$
\mathcal{A}_n \equiv \left(\frac{1}{k_\beta}n\hat{\beta}_n'\hat{\Sigma}_{\beta,\beta,n}^{-1}\hat{\beta}_n\right)^{1/2} \tag{A.6}
$$

where $\hat{\Sigma}_{\beta,\beta,n}$ is the upper $(p+1)\times(p+1)$ block of $\hat{\Sigma}_n$.

We use the following notation. $[z]$ rounds z to the nearest integer. $I(\cdot)$ is the indicator function: $I(A) = 1$ if A is true, otherwise $I(A) = 0$. $a_n/b_n \sim c$ implies $a_n/b_n \to c$ as $n \to \infty$. $|\cdot|$ is the l_1-matrix norm; $||\cdot||$ is the Euclidean norm; $||\cdot||_p$ is the L_p-norm. $K > 0$ is a finite constant whose value may change from place to place. $0_{a\times b}$ is an $a \times b$ dimensional matrix of zeros. *a.e.* denotes *almost everywhere*. $\Rightarrow^*$ denotes weak convergence on l_∞, the space of bounded functions with sup-norm topology, in the sense of Hoffman-Jϕrgensen (1984, 1991), cf. Dudley (1978) and Pollard (1984, 1990).

Recall that by *probability subadditivity*, for stochastic measurable $(\mathcal{A},\mathcal{B}) \geq 0$ and any $a \in$

$(0, \infty)$:

$$P(\mathcal{A} + \mathcal{B} > a) \leq P(\mathcal{A} > a/2) + P(\mathcal{B} > a/2). \tag{A.7}$$

Assumption 1 (data generating process, test weight)**.**

a. Identification*:*

(i) Under H_0, $E[\epsilon_t|x_t] = 0$ a.s. and $E[\epsilon_t^2|x_t] = \sigma_0^2$ a.s., a finite positive constant.

(ii) Under $\mathcal{C}(i, b)$: $E[(y_t - \zeta_0' x_t) d_{\psi,t}(\pi)] = 0$ for unique $\psi_0 = [0_{k_\beta}', \zeta_0']'$ in the interior of Ψ^*. Under $\mathcal{C}(ii, \omega_0)$: $E[\epsilon_t(\theta_0) \times d_{\theta,t}(\omega_0, \pi_0)] = 0$ for unique $\theta_0 = [\beta_0', \zeta_0', \pi_0']'$ in the interior of $\Theta^* = \Psi^* \times \Pi^*$.

b. Memory and Moments*: $\{\epsilon_t, x_t\}$ are L_p-bounded for some $p > 6$, strictly stationary, and β-mixing with mixing coefficients $\beta_l = O(l^{-qp(q-p)-\iota})$ for some $q > p$ and tiny $\iota > 0$.*

c. Response $g(x, \pi)$ and Test Weight $F(\lambda' \mathcal{W}(x))$*:*

(i) $g(\cdot, \pi)$ is Borel measurable for each π; $g(\cdot, \pi)$ is twice continuously differentiable in $\pi \in \mathbb{R}^{k_\pi}$; $g(x_t, \pi)$ is a non-degenerate random variable for each $\pi \in \Pi$.

(ii) $F : \mathbb{R} \to \mathbb{R}$ is analytic, non-polynomial, and $\mathcal{W}$ is one-to-one and bounded.

(iii) $E[\sup_{\pi \in \Pi} |(\partial/\partial\pi)^i g(x_t, \pi)|^6] < \infty$ and $E[\sup_{\lambda \in \Lambda} |(\partial/\partial\lambda)^j F(\lambda' \mathcal{W}(x_t))|^6] < \infty$ for $i = 0, 1, 2$ and $j = 0, 1$.

d. Long-Run Variances*:*

(i) Under $\mathcal{C}(i, b)$ with $||b|| < \infty$ let $\liminf_{n\to\infty} E[\inf_{\alpha,r,\theta} (r' \sum_{i=1}^m \alpha_i \mathcal{G}_{\psi,n}(\theta_i))^2] > 0$ and $\limsup_{n\to\infty} E[\sup_{\alpha,r,\theta} (r' \sum_{i=1}^m \alpha_i \mathcal{G}_{\psi,n}(\theta_i))^2] < \infty$.

(ii) Under $\mathcal{C}(ii, \omega_0)$ let $\liminf_{n\to\infty} E[\inf_{\alpha,r,\theta} (r' \sum_{i=1}^m \alpha_i \mathcal{G}_{\theta,n}(\theta_i))^2] > 0$ and $\limsup_{n\to\infty} E[\sup_{\alpha,r,\theta} (r' \sum_{i=1}^m \alpha_i \mathcal{G}_{\theta,n}(\theta_i))^2] < \infty$.

(iii) $E[\inf_{r,\omega,\pi} (r' d_{\theta,t}(\omega, \pi))^2] > 0$ and $E[\sup_{r,\omega,\pi} (r' d_{\theta,t}(\omega, \pi))^2] < \infty$; $E[\inf_{r,\pi} (r' d_{\psi,t}(\pi))^2] > 0$ and $E\left[\sup_{r,\pi} (r' d_{\psi,t}(\pi))^2\right] < \infty$.

(iv) $\liminf_{n\to\infty} \inf_{a,r,\pi} E[\mathbb{E}_{\psi,n}(\pi; a, r)^2] > 0$ and $\limsup_{n\to\infty} \sup_{a,r,\pi} E[\mathbb{E}_{\psi,n}(\pi; a, r)^2] < \infty$; and $\liminf_{n\to\infty} \inf_{a,r,\omega,\pi} E[\mathbb{E}_{\theta,n}(\omega, \pi; a, r)^2] > 0$ and $\limsup_{n\to\infty} \sup_{a,r,\omega,\pi} E[\mathbb{E}_{\theta,n}(\omega, \pi; a, r)^2] < \infty$.

(v) Under $\mathcal{C}(i, b)$ with $||b|| < \infty$, $\liminf_{n\to\infty} E[\sup_{\alpha,r,\lambda} \mathfrak{E}\mathfrak{G}_{\psi,n}(\lambda; a, r)^2] < \infty$.

(vi) Under $\mathcal{C}(ii, \omega_0)$, $E[\sup_{\alpha,r,\lambda} (1/\sqrt{n} \sum_{t=1}^n \epsilon_t \mathcal{K}_{\theta,t}(\lambda; a, m))^2] < \infty$ for each m.

e. True Parameter Space*:*

(i) $\Theta^* \equiv \{(\beta, \zeta, \pi) : \beta \in \mathcal{B}^*, \zeta \in \mathcal{Z}^*(\beta), \pi \in \Pi^*\}$ is compact.

(ii) $0_{k_\beta} \in \text{int}(\mathcal{B}^*)$.

(iii) For some set $\mathcal{Z}_0^*$ and some $\delta > 0$, $\mathcal{Z}^*(\beta) = \mathcal{Z}_0^*$ $\forall ||\beta|| < \delta$.

f. Optimization Parameter Space*:*

(i) $\Theta \equiv \{(\beta, \zeta, \pi) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\}$ and $\Theta^* \subset \text{int}(\Theta)$.

(*ii*) $(\Theta, \mathcal{B}, \Pi)$ *are compact, and* $\mathcal{Z}(\beta)$ *is compact for each* β. (*iii*) *For some set* $\mathcal{Z}_0$ *and some* $\delta > 0$, $\mathcal{Z}(\beta) = \mathcal{Z}_0 \; \forall ||\beta|| < \delta$ *and* $\mathcal{Z}_0^* \subset \text{int}(\mathcal{Z}_0)$.

Assumption 2 (identification of π)**.** *Let drift case* $\mathcal{C}(i, b)$ *hold with* $||b|| < \infty$. (*a*) *Each sample path of the process* $\{\xi_\psi(\pi, b) : \pi \in \Pi\}$ *in some set* $\mathfrak{A}(b)$ *with* $P(\mathfrak{A}(b)) = 1$ *is minimized over* Π *at a unique point* $\pi^*(b)$ *that may depend on the sample path.* (*b*) $P(\tau_\beta(\pi^*(b), b) = 0) = 0$.

Assumption 3 (non-degenerate scale on Λ-*a.e.*)**.**

a. Let $\mathcal{C}(i, b)$ *with* $||b|| < \infty$ *hold. Then* $P(E[\inf_{\pi \in \Pi}\{\epsilon_t^2(\psi_0, \pi)\}|x_t] > 0) = 1$. *There exists a Borel measurable function* $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$ *such that* $\kappa_t(\omega, \pi) \equiv [\mu(x_t), d_{\theta,t}(\omega, \pi)']'$ *has nonsingular* $E[\kappa_t(\omega, \pi)\kappa_t(\omega, \pi)']$ *uniformly on* $\{\omega \in \mathbb{R}^{k_x} : \omega'\omega = 1\} \times \Pi$.

b. Let $\mathcal{C}(ii, \omega_0)$ *hold. Then* $P(E[\epsilon_t^2|x_t] > 0) = 1$. *There exists a Borel measurable function* $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$ *such that* $\kappa_t \equiv [\mu(x_t), d_{\theta,t}]'$ *has a nonsingular* $E[\kappa_t\kappa_t']$.

Recall

$$\theta^+ \in \Theta^+ \equiv \left\{\theta^+ \in \mathbb{R}^{k_\beta + k_x + k_\pi + 1} : \theta^+ = [\|\beta\|, \omega(\beta), \zeta, \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\right\},$$

and

$$\begin{aligned}
&\epsilon_t(\theta^+) \equiv y_t - \zeta' x_t - ||\beta||\omega(\beta)' g(x_t, \pi) \\
&v^2(\theta^+, \lambda) = E\left[\epsilon_t^2(\theta^+)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\omega, \pi, \lambda)'\mathcal{H}_\theta^{-1}(\pi) d_{\theta,t}(\pi)\right\}^2\right]
\end{aligned}$$

and

$$\begin{aligned}
&\epsilon_t(\theta) \equiv y_t - \zeta' x_t - \beta' g(x_t, \pi) \\
&v^2(\theta, \lambda) = E\left[\epsilon_t^2(\theta)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\omega(\beta), \pi, \lambda)'\mathcal{H}_\theta^{-1}(\pi) d_{\theta,t}(\pi)\right\}^2\right].
\end{aligned}$$

Assumption 4 (non-degenerate scale)**.**

a. Let β *be a scalar. Let* $\inf_{\pi \in \Pi} v^2((\beta_0, \zeta_0, \pi), \lambda) > 0 \; \forall \lambda \in \Lambda$ *under identification case* $\mathcal{C}(i, b)$ *with* $|b| < \infty$, *and under* $\mathcal{C}(ii, \omega_0)$ *let* $v^2(\theta_0, \lambda) > 0 \; \forall \lambda \in \Lambda$.

b. Let β *be a vector. Let* $\inf_{\omega \in \mathbb{R}^{k_\beta} : \omega'\omega = 1, \pi \in \Pi} v^2((||\beta_0||, \omega, \zeta_0, \pi), \lambda) > 0 \; \forall \lambda \in \Lambda$ *under identification case* $\mathcal{C}(i, b)$ *with* $||b|| < \infty$, *and under* $\mathcal{C}(ii, \omega_0)$ *let* $v^2(\theta_0^+, \lambda) > 0 \; \forall \lambda \in \Lambda$.

Assumption 5 (p-value)**.** *a.* $\mathcal{F}_{\lambda,h}(c)$ *is continuous* a.e. *on* $[0, \infty)$, $\forall h \in \mathfrak{H}$. *b. The ICS-1 threshold sequence* $\{\kappa_n\}$ *satisfies* $\kappa_n \to \infty$ *and* $\kappa_n = o(\sqrt{n})$.

We exploit properties of the Vapnik-Červonenkis *subgraph* class of functions, denoted $\mathcal{V}(\mathcal{C})$. The $\mathcal{V}(\mathcal{C})$ class is large: it contains indicator, monotonic and continuous functions; and $\mathcal{V}(\mathcal{C})$

mappings of $\mathcal{V}(\mathcal{C})$ functions are in $\mathcal{V}(\mathcal{C})$, including linear combinations, minima, maxima, products and indicator transforms. See, e.g., van der Vaart and Wellner (1996, Chap. 2.6) for a compendium of $\mathcal{V}(\mathcal{C})$ properties.[1] See Vapnik and Červonenkis (1971), Dudley (1978, Section 7) and van der Vaart and Wellner (1996, Section 2), and see Pollard (1984, Chap. II.4) for the closely related *polynomial discrimination* class.

Assumption 6. *The test weight* $\{F(w) : w \in \mathbb{R}\}$ *and distribution functions* $\{F_{n,\lambda}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ *and* $\{F^*_{n,\lambda,h}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ *belong to the* $\mathcal{V}(\mathcal{C})$ *class.*

B Supporting Lemmata

All subsequent Gaussian processes have *almost surely* uniformly continuous and bounded sample paths, hence in many cases we just say *Gaussian process*. Let $\underline{\iota}(A)$ and $\bar{\iota}(A)$ denote the minimum and maximum eigenvalue of matrix A.

Lemma B.1. *Under* $\mathcal{C}(i,b)$ *and Assumption 1,* $\{\mathcal{G}_{\psi,n}(\theta) : \theta \in \Theta\} \Rightarrow^* \{\mathcal{G}_{\psi}(\theta) : \theta \in \Theta\}$*, a zero mean Gaussian process with* almost surely *uniformly continuous and bounded sample paths and covariance* $E[\mathcal{G}_{\psi}(\theta)\mathcal{G}_{\psi}(\tilde{\theta})']$, $||E[\mathcal{G}_{\psi}(\theta)\mathcal{G}_{\psi}(\theta)']|| < \infty$.

Proof. Recall Θ is compact and therefore bounded. Weak convergences to a Gaussian process with *almost surely* uniformly continuous and bounded sample paths therefore requires convergence in finite dimensional distributions, and stochastic equicontinuity (see, e.g., Dudley, 1978; Pollard, 1990).

Let $m \in \mathbb{N}$, $\alpha \in \mathbb{R}^m$ and $r \in \mathbb{R}^{k_x+k_\beta}$ be arbitrary, with $\alpha'\alpha = 1$ and $r'r = 1$. Under Assumption 1.b,c $\sum_{i=1}^m \alpha_i \epsilon_t(\theta_i) r' d_{\psi,t}(\theta_i)$ is, for any m-tuple $\{\theta_1, ..., \theta_m\}$ of points θ_i in Θ, strictly stationary, L_p-bounded, $p > 4$, and β-mixing with coefficients $\beta_l = O(l^{-(pq/(q-p))-\iota})$ for some $\iota > 0$ and $q > p$. Hence $E[(\sum_{i=1}^m \alpha_i r' \mathcal{G}_{\psi,n}(\theta_i))^2] = O(1)$ (McLeish, 1975, Theorem 1.6, Lemma 2.1). Long run variance Assumption 1.d(i) and Theorem 1.4 in Ibragimov (1962) therefore yield: $\sum_{i=1}^m \alpha_i r' \mathcal{G}_{\psi,n}(\theta_i) \overset{d}{\to} N(0, \lim_{n\to\infty} E[(\sum_{i=1}^m \alpha_i r' \mathcal{G}_{\psi,n}(\theta_i))^2])$. Convergence in finite dimensional distributions now follows from the Cramér-Wold theorem.

Stochastic equicontinuity for $r'\mathcal{G}_{\psi,n}(\theta)$ holds if $\forall(\epsilon, \eta) > 0$ there exists $\delta > 0$ such that:

$$\lim_{n\to\infty} \mathcal{P}_n(r, \delta, \eta) = \lim_{n\to\infty} P\left(\sup_{\theta,\tilde{\theta}\in\Theta:||\theta-\tilde{\theta}||\leq\delta} \left|r'\mathcal{G}_{\psi,n}(\theta) - r'\mathcal{G}_{\psi,n}(\tilde{\theta})\right| > \eta\right) < \epsilon. \tag{B.8}$$

We adapt arguments developed in Arcones and Yu (1994, proof of Theorem 2.1 and Lemma 2.1) to prove (B.8). This requires the $\mathcal{V}(\mathcal{C})$ *subgraph* class of functions. By the implication of

[1] We exploit the facts that an indicator function of a $\mathcal{V}(\mathcal{C})$ index function is in $\mathcal{V}(\mathcal{C})$, and a continuous function evaluated at a $\mathcal{V}(\mathcal{C})$ function is in $\mathcal{V}(\mathcal{C})$.

probability subadditivity (A.7) and $r'r = 1$, it suffices to prove the claim for each element of $\mathcal{G}_{\psi,n}(\theta) = [\mathcal{G}_{\psi,n,i}(\theta)]_{i=1}^{k_x+k_\beta}$.

$\mathcal{G}_{\psi,n,i}(\theta)$ lies in $\mathcal{V}(\mathcal{C})$ because it is continuous, hence the covering numbers satisfy $\mathcal{N}(\varepsilon, \mathcal{K}, ||\cdot||_2)$ $< a\varepsilon^{-b}$ for all $\varepsilon \in (0,1)$ and some $a, b > 0$ (e.g. Lemma 7.13 in Dudley, 1978, and Lemma II.25 in Pollard, 1984). Furthermore, under Assumption 1.b,c each $\mathcal{G}_{\psi,n,i}(\theta)$ is L_r-bounded, $r \equiv p/2$ > 2, and β-mixing with coefficients $\beta_l = O(l^{-qp/(q-p)-\iota})$, $q > p > 6$ and tiny $\iota > 0$. By simple algebra it follows $\beta_l = O(l^{-r/(r-2)}) = O(l^{-p/(p-4)})$ because $p/(p-4) < qp/(q-p)$. Therefore $\{\mathcal{G}_{\psi,n,i}(\theta) : \theta \in \Theta\}$ is stochastically equicontinuous by Lemma 2.1 in Arcones and Yu (1994, see especially the argument following eq. (2.13)). $\mathcal{QED}$

Lemma B.2. *Under* $\mathcal{C}(i,b)$ *and Assumption 1,* $\sup_{\pi\in\Pi} ||\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)|| \xrightarrow{p} 0$*, where* $\underline{\iota}(\mathcal{H}_\psi(\pi))$ > 0 *and* $\bar{\iota}(\mathcal{H}_\psi(\pi)) < \infty$ *for each* $\pi \in \Pi$.

Proof. We have $\widehat{\mathcal{H}}_{\psi,n}(\pi) \xrightarrow{p} \mathcal{H}_\psi(\pi)$ pointwise under Assumption 1.b,c since $d_{\psi,t}(\kappa)$ is stationary, L_2-bounded, and ergodic by the β-mixing property. Further, $\underline{\iota}(\mathcal{H}_\psi(\pi)) > 0$ and $\bar{\iota}(\mathcal{H}_\psi(\pi)) <$ ∞ for each $\pi \in \Pi$ respectively follow from $\inf_{r'r=1} E\left[(r'd_{\psi,t}(\pi))^2\right] > 0$ under Assumption 1.d(iii), and $||\mathcal{H}_\psi(\pi)|| < \infty$ under envelope bounds Assumption 1.c and compactness of Θ.

It remains to show $\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)$ is stochastically equicontinuous. By the mean-value-theorem and Cauchy-Schwartz inequality:

$$
\begin{aligned}
&E\left[\sup_{\pi,\tilde{\pi}\in\Pi:||\pi-\tilde{\pi}||\leq\delta}\left|\widehat{\mathcal{H}}_{\psi,n}(\pi) - \widehat{\mathcal{H}}_{\psi,n}(\tilde{\pi})\right|\right] \\
&\qquad \leq 2E\left[\sup_{\pi\in\Pi}\left|\frac{\partial}{\partial\pi}d_{\psi,t}(\pi)\right|\sup_{\pi\in\Pi}|d_{\psi,t}(\pi)'|\right]\times\delta \\
&\qquad \leq 2\left(E\left[\sup_{\pi\in\Pi}\left|\frac{\partial}{\partial\pi}g(x_t,\pi)\right|^2\right]\right)^{1/2}\left(E\left[\left(\sup_{\pi\in\Pi}|g(x_t,\pi)| + |x_t|\right)^2\right]\right)^{1/2}\times\delta \equiv \mathcal{K}\delta,
\end{aligned}
$$

where $\mathcal{K} \geq 0$ is implicitly defined and $\delta > 0$. The right hand side is bounded by L_2-boundedness of x_t, $\sup_{\pi\in\Pi}|g(x_t,\pi)|$ and $\sup_{\pi\in\Pi}|(\partial/\partial\pi)g(x_t,\pi)|$ under Assumption 1.b,c. Hence $\mathcal{K} \in [0,\infty)$. Therefore, assuming $\mathcal{K} > 0$, $\forall(\epsilon,\eta) > 0$ there exists δ, $0 < \delta < \epsilon/\mathcal{K}$, such that by Markov's inequality:

$$
\lim_{n\to\infty} P\left(\sup_{\pi,\tilde{\pi}\in\Pi:||\pi-\tilde{\pi}||\leq\delta}\left|\left\{\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)\right\} - \left\{\widehat{\mathcal{H}}_{\psi,n}(\tilde{\pi}) - \mathcal{H}_\psi(\tilde{\pi})\right\}\right| > \eta\right) < \epsilon. \tag{B.9}
$$

If $\mathcal{K} = 0$ then $\forall(\epsilon,\eta) > 0$ and any $\delta \in (0,\infty)$ (B.9) holds. This yields stochastic equicontinuity, completing the proof. $\mathcal{QED}$

Lemma B.3. *Under $\mathcal{C}(ii,\omega_0)$ and Assumption 1, $\{\mathcal{G}_{\theta,n}(\theta) : \theta \in \Theta\} \Rightarrow^* \{\mathcal{G}_\theta(\theta) : \theta \in \Theta\}$, a zero mean Gaussian process with* almost surely *uniformly continuous and bounded sample paths.*

Proof. The arguments used to prove Lemma B.1 carry over verbatim, except long run variance Assumption 1.d(ii) is used in place of Assumption 1.d(i). $\mathcal{QED}$.

Corollary B.4. *Let $\theta_n \equiv [\beta_n', \zeta_0', \pi_0']'$ be the sequence of true values under local drift $\{\beta_n\}$. Under $\mathcal{C}(ii,\omega_0)$ and Assumption 1, $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial\theta)Q_n(\theta_n) \stackrel{d}{\rightarrow} \mathcal{G}_\theta$, a zero mean Gaussian law with a finite, positive definite covariance $E[\mathcal{G}_\theta\mathcal{G}_\theta']$, and has a version that has* almost surely *uniformly continuous and bounded sample paths. Moreover, $E[\mathcal{G}_\theta\mathcal{G}_\theta'] = \sigma^2 E[d_{\theta,t}d_{\theta,t}']$ under H_0.*

Proof. By the definition of $\mathcal{G}_{\theta,n}(\theta_n)$:

$$\sqrt{n}\mathfrak{B}(\beta_n)^{-1}\frac{\partial}{\partial\theta}Q_n(\theta_n) = \mathcal{G}_{\theta,n}(\theta_n) + \sqrt{n}E\left[\epsilon_t(\theta_n)d_{\theta,t}\left(\beta_n/\left\|\beta_n\right\|,\pi_0\right)\right].$$

Combine Lemma B.3, $\theta_n \rightarrow \theta_0$, the fact that θ_n is non-random, and continuity to yield $\mathcal{G}_{\theta,n}(\theta_n) \stackrel{d}{\rightarrow} \mathcal{G}_\theta \equiv \mathcal{G}_\theta(\theta_0)$. By identification Assumption 1.a(ii) and the fact that $\theta_n \equiv [\beta_n', \zeta_0', \pi_0']'$ is the sequence of true values under local drift $\{\beta_n\}$, it follows that $E[\epsilon_t(\theta_n)d_{\theta,t}(\beta_n/||\beta_n||,\pi_0)] = 0$. This proves $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial\theta)Q_n(\theta_n) \stackrel{d}{\rightarrow} \mathcal{G}_\theta$.

Finally, since $\theta_n \equiv [\beta_n', \zeta_0', \pi_0']'$ is the sequence of true values, under H_0 note that

$$\begin{aligned}\mathcal{G}_{\theta,n}(\theta_n) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t d_{\theta,t}(\omega(\beta_n),\pi) - E\left[\epsilon_t(\theta_n)d_{\theta,t}(\omega(\beta_n),\pi)\right]\right\} \\ &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\theta,t}(\omega(\beta_n),\pi).\end{aligned}$$

Hence, in view of stationarity:

$$E[\mathcal{G}_{\theta,n}(\theta_n)\mathcal{G}_{\theta,n}(\theta_n)'] = \sigma^2 E\left[d_{\theta,t}\left(\beta_n/\left\|\beta_n\right\|,\pi_0\right)d_{\theta,t}(\beta_n/\left\|\beta_n\right\|,\pi_0)\right]'.$$

Since $\beta_n/||\beta_n|| \rightarrow \omega_0$ and $||\omega_0|| = 1$, under Assumption 1.b,c:

$$\mathfrak{d}_t \equiv \limsup_{n\rightarrow\infty}\sup_{r'r=1}\left(r'\left[g(x_t,\pi)', x_t', \frac{\beta_n'}{\|\beta_n\|}\frac{\partial}{\partial\pi}g(x_t,\pi)\right]'\right)^2$$

exists and $E[\mathfrak{d}_t] < \infty$. Dominated convergence now yields

$$E\left[d_{\theta,t}\left(\beta_n/\left\|\beta_n\right\|,\pi_0\right)d_{\theta,t}(\beta_n/\left\|\beta_n\right\|,\pi_0)\right] \rightarrow E[d_{\theta,t}d_{\theta,t}'],$$

hence $E[\mathcal{G}_{\theta,n}(\theta_n)\mathcal{G}_{\theta,n}(\theta_n)'] \rightarrow \sigma^2 E[d_{\theta,t}d_{\theta,t}']$.This implies $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial\theta)Q_n(\theta_n) \stackrel{d}{\rightarrow} \mathcal{G}_\theta$, with asymptotic variance $\sigma^2 E[d_{\theta,t}d_{\theta,t}']$ as required. $\mathcal{QED}$

Lemma B.5. *Under $\mathcal{C}(ii,\omega_0)$ and Assumption 1, $\widehat{\mathcal{H}}_n \stackrel{p}{\rightarrow} \mathcal{H}_\theta$, and $\underline{\iota}(\mathcal{H}_\theta) > 0$ and $\bar{\iota}(\mathcal{H}_\theta) < \infty$.*

Proof. By the construction of $\widehat{\mathcal{H}}_n \equiv 1/n\sum_{t=1}^n d_{\theta,t}(\omega(\beta_n),\pi_0)d_{\theta,t}(\omega(\beta_n),\pi_0)'$ and $\mathcal{H}_\theta \equiv E[d_{\theta,t}d_{\theta,t}']$, and $d_{\theta,t}(\omega,\pi) \equiv [g(x_t,\pi)', x_t', \omega'(\partial/\partial\pi)g(x_t,\pi)]'$, after adding and subtracting like terms, we have for any $r = [r_\beta', r_x', r_\pi']'$, $r_\beta \in \mathbb{R}^{k_\beta}, r_x \in \mathbb{R}^{k_x}, r_\pi \in \mathbb{R}^{k_\pi}$:

$$
\begin{aligned}
r'\left(\widehat{\mathcal{H}}_n - \mathcal{H}_\theta\right)r \;=\; & \frac{1}{n}\sum_{t=1}^n \left(r_\beta' g(x_t,\pi_0) + r_x' x_t + r_\pi' \frac{\partial}{\partial\pi'}g(x_t,\pi_0)\omega_0\right)^2 \\
& -E\left[\left(r_\beta' g(x_t,\pi_0) + r_x' x_t + r_\pi' \frac{\partial}{\partial\pi'}g(x_t,\pi_0)\omega_0\right)^2\right] \\
& +\frac{1}{n}\sum_{t=1}^n \left(r_\pi' \frac{\partial}{\partial\pi'}g(x_t,\pi_0)\left(\frac{\beta_n}{\|\beta_n\|} - \omega_0\right)\right)^2 \\
& +2\frac{1}{n}\sum_{t=1}^n \left(r_\beta' g(x_t,\pi_0) + r_x' x_t + r_\pi' \frac{\partial}{\partial\pi'}g(x_t,\pi_0)\omega_0\right) \times r_\pi' \frac{\partial}{\partial\pi'}g(x_t,\pi_0)\left(\frac{\beta_n}{\|\beta_n\|} - \omega_0\right).
\end{aligned}
$$

The Assumption 1.b,c envelop moment and mixing properties imply each summand is a summation of stationary, ergodic and integrable random variables. Further $\beta_n/||\beta_n|| - \omega_0 \rightarrow 0$ by assumption. The ergodic theorem now yields $r'(\widehat{\mathcal{H}}_n - \mathcal{H}_\theta)r \stackrel{p}{\rightarrow} 0$.

Finally, $\underline{\iota}(\mathcal{H}_\theta) > 0$ and $\bar{\iota}(\mathcal{H}_\theta) < \infty$ follow from Assumption 1.c,d(iii). $\mathcal{QED}$

Define the augmented parameter, and its space:

$$
\begin{aligned}
\theta^+ \;&\equiv\; [\|\beta\|, \omega', \zeta', \pi']' \\
&\in\; \Theta^+ \equiv \left\{\theta^+ \in \mathbb{R}^{k_x+k_\beta+k_\pi+1} : \theta^+ = [\|\beta\|, \omega(\beta), \zeta, \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\right\}.
\end{aligned}
$$

Define

$$\epsilon_t(\theta^+) \equiv y_t - \zeta' x_t - \|\beta\|\,|\omega' g(x_t,\pi),$$

and:

$$\widehat{\mathcal{H}}_n(\theta^+) \equiv \frac{1}{n}\sum_{t=1}^n d_{\theta,t}(\omega(\beta),\pi)d_{\theta,t}(\omega(\beta),\pi)', \;\; \hat{\mathcal{V}}_n(\theta^+) \equiv \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta^+)d_{\theta,t}(\omega(\beta),\pi)d_{\theta,t}(\omega(\beta),\pi)'.$$

Hence $\widehat{\mathcal{H}}_n(\hat{\theta}_n^+) = \widehat{\mathcal{H}}_n$ and $\hat{\mathcal{V}}_n(\hat{\theta}_n^+) = \hat{\mathcal{V}}_n$. Define

$$\mathcal{H}_\theta(\theta^+) \equiv E\left[d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)'\right] \text{ and } \mathcal{V}(\theta^+) \equiv E\left[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)'\right].$$

In the interest of decreasing (some) notation we use the same argument θ^+ for both $\widehat{\mathcal{H}}_n(\theta^+)$ and $\hat{\mathcal{V}}_n(\theta^+)$, although $\widehat{\mathcal{H}}_n(\theta^+)$ only depends on $(\omega(\beta), \pi)$.

Lemma B.6. *Under Assumption 1,* $\sup_{\theta^+ \in \Theta^+:} ||\widehat{\mathcal{H}}_n(\theta^+) - \mathcal{H}_\theta(\theta^+)|| \xrightarrow{p} 0$*,* $\sup_{\pi \in \Pi} ||\hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0) - \mathcal{D}_\psi(\pi)|| \xrightarrow{p} 0$*, and* $\sup_{\theta^+ \in \Theta^+:} ||\hat{\mathcal{V}}_n(\theta^+) - \mathcal{V}(\theta^+)|| \xrightarrow{p} 0$*, where* $\inf_{\theta^+ \in \Theta^+:} \underline{\iota}(\mathcal{H}_\theta(\theta^+)) > 0$*,* $\bar{\iota}(\mathcal{H}_\theta) < \infty$*,* $\inf_{\theta^+ \in \Theta^+:} \underline{\iota}(\mathcal{V}(\theta^+)) > 0$*, and* $\bar{\iota}(\mathcal{V}_\theta) < \infty$*.*

Proof. We prove the claim for $\hat{\mathcal{V}}_n(\theta^+)$, the proofs for $\widehat{\mathcal{H}}_n(\theta^+)$ and $\hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0)$ being similar. Pointwise convergence follows from mixing (hence ergodicity) and moment properties in Assumption 1.b,c.

Uniform convergence is proven if we show stochastic equicontinuity: $\forall(\epsilon, \eta) > 0$ there exists $\delta > 0$ such that:

$$\begin{aligned}
&\lim_{n\to\infty} \mathcal{P}_n(r, \delta, \eta) \qquad\qquad (B.10)\\
&\quad = \lim_{n\to\infty} P\left(\sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left| \left\{\hat{\mathcal{V}}_n(\theta^+) - \mathcal{V}(\theta^+)\right\} - \left\{\hat{\mathcal{V}}_n(\tilde{\theta}) - \mathcal{V}(\tilde{\theta}^+)\right\}\right| > \eta\right)\\
&\quad < \epsilon.
\end{aligned}$$

First note that:

$$\begin{aligned}
&E\left[\sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left|\hat{\mathcal{V}}_n(\theta^+) - \hat{\mathcal{V}}_n(\tilde{\theta})\right|\right]\\
&\quad = \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left| \frac{1}{n} \sum_{t=1}^n \left\{ \epsilon_t^2(\theta^+) d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' - \epsilon_t^2(\tilde{\theta}^+) d_{\theta,t}(\tilde{\omega}, \tilde{\pi}) d_{\theta,t}(\tilde{\omega}, \tilde{\pi})' \right\} \right|\\
&\quad \le \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left| \frac{1}{n} \sum_{t=1}^n \left\{ \epsilon_t^2(\theta^+) - \epsilon_t^2(\tilde{\theta}^+) \right\} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \right|\\
&\qquad + \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left| \frac{1}{n} \sum_{t=1}^n \epsilon_t^2(\tilde{\theta}^+) \left\{ d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' - d_{\theta,t}(\tilde{\omega}, \tilde{\pi}) d_{\theta,t}(\tilde{\omega}, \tilde{\pi})' \right\} \right|.
\end{aligned}$$

By the mean value theorem, and the moment properties of Assumption 1.b,c:

$$\begin{aligned}
&E\left[\sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: ||\theta^+ - \tilde{\theta}^+|| \le \delta} \left| \frac{1}{n} \sum_{t=1}^n \left\{ \epsilon_t^2(\theta^+) - \epsilon_t^2(\tilde{\theta}^+) \right\} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \right|\right]\\
&\qquad \le 2E\left[\sup_{\theta^+ \in \Theta^+} \left|\epsilon_t(\theta^+)\right| \sup_{\theta^+ \in \Theta^+} \left|d_{\theta,t}(\omega, \pi)\right|^3\right] \times \delta \le K\delta,
\end{aligned}$$

and

$$E\left[\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\tilde{\theta}^+)\left\{d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)'-d_{\theta,t}(\tilde{\omega},\tilde{\pi})d_{\theta,t}(\tilde{\omega},\tilde{\pi})'\right\}\right|\right]$$
$$\leq 2E\left[\sup_{\theta^+\in\Theta^+}\left|\epsilon_t^2(\theta^+)\right|\sup_{\theta^+\in\Theta^+}\left|d_{\theta,t}(\omega,\pi)\right|\sup_{\theta^+\in\Theta^+}\left|\frac{\partial}{\partial\theta^+}d_{\theta,t}(\omega,\pi)\right|\right]\times\delta\leq K\delta,$$

where

$$\left|\frac{\partial}{\partial\theta^+}d_{\theta,t}(\omega,\pi)\right|\leq 2\times\left|\frac{\partial}{\partial\pi}g(x_t,\pi)\right|+|\omega|\times\left|\frac{\partial^2}{\partial\pi\partial\pi'}g(x_t,\pi)\right|.$$

A similar set of steps shows

$$\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|\mathcal{V}_n(\theta^+)-\mathcal{V}_n(\tilde{\theta})\right|$$
$$=\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|E\left[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi_0)d_{\theta,t}(\omega,\pi_0)'\right]-E\left[\epsilon_t^2(\tilde{\theta}^+)d_{\theta,t}(\tilde{\omega},\tilde{\pi})d_{\theta,t}(\tilde{\omega},\tilde{\pi})'\right]\right|$$
$$\leq K\delta.$$

Now invoke Markov and Minkowski inequalities to yield:

$$\lim_{n\to\infty}P\left(\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|\left\{\hat{\mathcal{V}}_n(\theta^+)-\mathcal{V}(\theta^+)\right\}-\left\{\hat{\mathcal{V}}_n(\tilde{\theta})-\mathcal{V}(\tilde{\theta}^+)\right\}\right|>\eta\right)$$
$$\leq\lim_{n\to\infty}\frac{1}{\eta}E\left[\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|\hat{\mathcal{V}}_n(\theta^+)-\hat{\mathcal{V}}_n(\tilde{\theta})\right|\right]$$
$$+\lim_{n\to\infty}\frac{1}{\eta}\sup_{\theta^+,\tilde{\theta}^+\in\Theta^+:||\theta^+-\tilde{\theta}^+||\leq\delta}\left|\left\{\mathcal{V}(\theta^+)\right\}-\mathcal{V}(\tilde{\theta}^+)\right|$$
$$\leq K\delta.$$

This proves stochastic equicontinuity (B.10) for any δ such that $0<\delta<\epsilon/K$. $\mathcal{QED}$

Define

$$a_n\equiv\begin{cases}\sqrt{n} & \text{if } \mathcal{C}(i,b) \text{ and } \|b\|<\infty\\ \|\beta_n\|^{-1} & \text{if } \mathcal{C}(i,b) \text{ and } \|b\|=\infty\end{cases}$$

Recall

$$\psi_{0,n}\equiv\left[0'_{k_\beta},\zeta'_0\right]',$$

hence $Q_{0,n}\equiv Q_n(\psi_{0,n},\pi)$ does not depend on π. Define:

$$\mathcal{Z}_n(\pi)=-a_n\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi).$$

Under $\mathcal{C}(i,b)$, Lemma B.2 yields that $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ is positive definite uniformly on Π, asymptotically with probability approaching one. Write $Q_n^c(\pi) \equiv Q_n(\hat{\psi}_n(\pi),\pi)$.

Lemma B.7. *Let drift case $\mathcal{C}(i,b)$ and Assumption 1 hold.*

a. In general $a_n(\hat{\psi}_n(\pi) - \psi_{0,n}) = \mathcal{Z}_n(\pi)$.

b. $a_n^2\{Q_n^c(\pi) - Q_{0,n}\} = -2^{-1}\mathcal{Z}_n(\pi)'\widehat{\mathcal{H}}_{\psi,n}(\pi)\mathcal{Z}_n(\pi)$ where $Q_{0,n} \equiv Q_n(\psi_{0,n},\pi)$.

Proof.

Claim a. By the definition of $\hat{\psi}_n(\pi)$, $0 = 1/n\sum_{t=1}^n \epsilon_t(\hat{\psi}_n(\pi),\pi)d_{\psi,t}(\pi)$. Now use $(\partial/\partial\psi)Q_n(\psi_{0,n},\pi)$ $= -1/n\sum_{t=1}^n \epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)$, $\widehat{\mathcal{H}}_{\psi,n}(\pi) \equiv 1/n\sum_{t=1}^n d_{\psi,t}(\pi)d_{\psi,t}(\pi)'$, and linearity of the first order equation in $\hat{\psi}_n(\pi)$, to yield the desired result.

Claim b. The equality $x^2 - y^2 = (x-y)(x+y)$ and rudimentary algebra yield:

$$\begin{aligned}
Q_n^c(\pi) - Q_{0,n} &= -\frac{1}{n}\sum_{t=1}^n \epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)' \times \left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) \\
&\quad + \frac{1}{2}\left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) \times \widehat{\mathcal{H}}_{\psi,n}(\pi) \times \left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) \\
&= -\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi)' \times \left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) + \frac{1}{2}\left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) \times \widehat{\mathcal{H}}_{\psi,n}(\pi) \times \left(\hat{\psi}_n(\pi) - \psi_{0,n}\right).
\end{aligned}$$

Use (a) and the form of $\mathcal{Z}_n(\pi)$ to deduce $a_n(\partial/\partial\psi)Q_n(\psi_{0,n},\pi)'\mathcal{Z}_n(\pi) = \mathcal{Z}_n(\pi)\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)\mathcal{Z}_n(\pi)$ hence:

$$\begin{aligned}
a_n^2\left\{Q_n^c(\pi) - Q_{0,n}\right\} &= -a_n\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi)'\mathcal{Z}_n(\pi) + \frac{1}{2}\mathcal{Z}_n(\pi)'\widehat{\mathcal{H}}_{\psi,n}(\pi)\mathcal{Z}_n(\pi) \\
&= -\frac{1}{2}\mathcal{Z}_n(\pi) \times \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \times \mathcal{Z}_n(\pi).
\end{aligned}$$

This proves the claim and completes the proof. $\mathcal{QED}$

Define

$$\vartheta_\psi(\pi,\omega_0) \equiv -2^{-2}\omega_0'\mathcal{D}_\psi(\pi)'\mathcal{H}_\psi^{-1}(\pi)\mathcal{D}_\psi(\pi)\omega_0$$

where

$$\mathcal{D}_\psi(\pi) = -E\left[d_{\psi,t}(\pi)g(x_t,\pi_0)'\right].$$

Recall from the main paper:

$$\xi_\psi(\pi,b) \equiv -\frac{1}{2}\left\{\mathcal{G}_\psi(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\}'\mathcal{H}_\psi^{-1}(\pi)\left\{\mathcal{G}_\psi(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\}.$$

The following is a key result for characterizing the asymptotic properties of $\hat{\pi}_n$ under weak identification.

Lemma B.8. *Let drift case $\mathcal{C}(i,b)$ and Assumption 1 hold.*

a. If $||b|| < \infty$ then $\{n(Q_n^c(\pi) - Q_{0,n}) : \pi \in \Pi\} \Rightarrow^* \{\xi_\psi(\pi,b) : \pi \in \Pi\}$.

b. If $||b|| = \infty$ and $\beta_n/||\beta_n|| \to \omega_0$ for some $\omega_0 \in \mathbb{R}^{k_\beta}$, $||\omega_0|| = 1$, then

$$\sup_{\pi\in\Pi}\left|\frac{1}{\|\beta_n\|^2}\left(Q_n^c(\pi) - Q_{0,n}\right) - \vartheta_\psi(\pi,\omega_0)\right| \xrightarrow{p} 0.$$

Proof.

Claim a. Recall

$$\mathcal{G}_{\psi,n}(\theta) = \sqrt{n}\left\{\frac{\partial}{\partial\psi}Q_n(\theta) + E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\}.$$

By Lemma B.7.b and $||b|| < \infty$:

$$\begin{aligned} n\left(Q_n^c(\pi),\pi) - Q_{0,n}\right) &= -n\frac{1}{2}\mathcal{Z}_n(\pi)'\widehat{\mathcal{H}}_{\psi,n}(\pi)\mathcal{Z}_n(\pi) = -\frac{1}{2}\sqrt{n}\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi)'\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)\sqrt{n}\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi) \\ &= -\frac{1}{2}\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) - \sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\}' \times \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \\ &\quad \times\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) - \sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\}. \end{aligned}$$

Further, by (C.18) in the proof of Theorem 4.1 in Appendix C:

$$\sup_{\pi\in\Pi}\left|\sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right] + \mathcal{D}_\psi(\pi)b\right| \to 0.$$

Now use Lemma B.1 for $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)$, and Lemma B.2 for $\widehat{\mathcal{H}}_{\psi,n}(\pi)$, to prove the claim.

Claim b. Lemma B.7.b and the definition of $\mathcal{Z}_n(\pi)$ lead to:

$$\begin{aligned} a_n^2\{Q_n^c(\pi) - Q_{0,n}\} = -\frac{1}{2}\frac{1}{\sqrt{n}\,\|\beta_n\|}&\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) - \sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\}'\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \\ &\times\frac{1}{\sqrt{n}\,\|\beta_n\|}\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) - \sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\}. \end{aligned}$$

By (C.17) in the proof of Theorem 4.1:

$$\sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right] = \sqrt{n}E\left[\{\epsilon_t(\psi_{0,n},\pi) - \epsilon_t(\theta_n)\}\,d_{\psi,t}(\pi)\right] = E\left[\sqrt{n}\beta_n'g(x_t,\pi_0)d_{\psi,t}(\pi)\right],$$

hence $||\beta_n||^{-1}E[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)] = E[||\beta_n||^{-1}\beta_n' g(x_t,\pi_0)d_{\psi,t}(\pi)]$, and therefore

$$\sup_{\pi\in\Pi}\left|\frac{1}{||\beta_n||}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]+\mathcal{D}_\psi(\pi)\omega_0\right|\to 0.$$

By supposition $\sqrt{n}||\beta_n|||\to\infty$, hence Lemma B.1 with the continuous mapping theorem, and Cramér's Theorem, yield:

$$\sup_{\pi\in\Pi}\left\|\frac{1}{\sqrt{n}\,\|\beta_n\|}\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)\right\|\leq\frac{1}{\inf_{\pi\in\Pi}\sqrt{n}\,\|\beta_n\|}\sup_{\pi\in\Pi}\|\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)\|\xrightarrow{p}0.$$

Lemma B.2 applied to $\widehat{\mathcal{H}}_{\psi,n}(\pi)$, and the Slutsky theorem complete the proof. $\mathcal{QED}$

Write $\epsilon_t(\psi,\pi)=y_t-\zeta' x_t-\beta' g(x_t,\pi)$. Recall ψ_n is the (possibly drifting) true value of $\psi=[\beta',\zeta']'$ under H_0.

Lemma B.9. *Let Assumption 1 hold.*

a. Under $\mathcal{C}(i,b)$ with $||b||<\infty$:

$$\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)-E\left[\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)\right]\right\}:\Pi,\Lambda\right\}\Rightarrow^*\left\{\mathfrak{Z}_\psi(\pi,\lambda):\Pi,\Lambda\right\},$$

a zero mean Gaussian process with covariance kernel $E[\mathfrak{Z}_\psi(\pi,\lambda)\mathfrak{Z}_\psi(\tilde{\pi},\tilde{\lambda})]$. Under H_0,

$$\sup_{\pi\in\Pi,\lambda\in\Lambda}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)-E\left[\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)\right]\right\}-\frac{1}{\sqrt{n}}\sum_{t=1}^n\epsilon_t\mathcal{K}_{\psi,t}(\pi,\lambda)\right|\xrightarrow{p}0,\quad\text{(B.11)}$$

and $E[\mathfrak{Z}_\psi(\pi,\lambda)\mathfrak{Z}_\psi(\tilde{\pi},\tilde{\lambda})]=\sigma^2E[\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})]$.

b. Under $\mathcal{C}(i,\omega_0)$, $\{1/\sqrt{n}\sum_{t=1}^n\epsilon_t\mathcal{K}_{\theta,t}(\lambda):\lambda\in\Lambda\}\Rightarrow^*\{\mathfrak{Z}_\theta:\lambda\in\Lambda\}$, a zero mean Gaussian process with covariance $E[\mathfrak{Z}_\theta(\lambda)\mathfrak{Z}_\theta\tilde{(\lambda)}]=E[\epsilon_t^2\mathcal{K}_{\theta,t}(\lambda)\mathcal{K}_{\theta,t}(\tilde{\lambda})]$ where $\mathcal{K}_{\theta,t}(\lambda)\equiv F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}$.

Proof. We only prove Claim (a). The proof for Claim (b) is nearly identical.

Π,Λ are compact and therefore bounded. Weak convergences to a Gaussian process with *almost surely* uniformly continuous and bounded sample paths requires convergence in finite dimensional distributions, and stochastic equicontinuity (see, e.g., Dudley, 1978; Pollard, 1990).

Write compactly $\chi\equiv[\pi',\lambda']'\in\mathcal{X}\equiv\Pi\times\Lambda$, and define:

$$\mathcal{E}_{\psi,t}(\psi_n,\chi)\equiv\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)-E\left[\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)\right]$$

$$\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m) \equiv \sum_{i=1}^{m} \alpha_i \mathcal{E}_{\psi,t}(\psi_n, \chi_i)$$

where $m \in \mathbb{N}$, $a \in \mathbb{R}^m$ satisfies $a'a = 1$, and $\{\chi_1, ..., \chi_m\}$ is an m-tuple of points $\chi_i = [\pi_i', \lambda_i']' \in \mathcal{X}$. Under Assumption 1.b,c $\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m)$ has a zero mean, and is strictly stationary, L_p-bounded, $p > 4$, and β-mixing with coefficients $\beta_l = O(l^{-(pq/(q-p))-\iota})$ for some $\iota > 0$ and $q > p$. Hence $E[\{1/\sqrt{n}\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m)\}^2] = O(1)$ (McLeish, 1975, Theorem 1.6, Lemma 2.1). Long run variance Assumption 1.d(v) coupled with Assumption 4 imply $E[(\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m))^2] \to \infty$. Now invoke Theorem 1.4 in Ibragimov (1962) to yield:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m) \xrightarrow{d} N\left(0, \lim_{n\to\infty} E\left[\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m)\right\}^2\right]\right),$$

where $\lim_{n\to\infty} E[\{1/\sqrt{n}\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m)\}\}^2] < \infty$. Convergence in finite dimensional distributions now follows by the Cramér-Wold theorem.

Next, after adding and subtracting $\beta_n' g(x_t, \pi_0)$:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{E}_{\psi,t}(\psi_n, \chi) \\
&\quad= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t \mathcal{K}_{\psi,t}(\pi, \lambda) - E\left[\epsilon_t \mathcal{K}_{\psi,t}(\pi, \lambda)\right]\right\} \\
&\qquad - \sqrt{n}\beta_n' \frac{1}{n}\sum_{t=1}^{n}\left\{x_t\left\{g(x_t, \pi) - g(x_t, \pi_0)\right\}\mathcal{K}_{\psi,t}(\pi, \lambda) - E\left[x_t\left\{g(x_t, \pi) - g(x_t, \pi_0)\right\}\mathcal{K}_{\psi,t}(\pi, \lambda)\right]\right\} \\
&\quad= \mathfrak{Z}_n(\pi, \lambda) + \mathfrak{X}_n(\pi, \lambda).
\end{aligned}$$

Under H_0 and Assumption 1.a, $E[\epsilon_t \mathcal{K}_{\psi,t}(\pi, \lambda)] = 0$ and

$$E\left[\mathfrak{Z}_n(\pi, \lambda)\mathfrak{Z}_n(\tilde{\pi}, \tilde{\lambda})\right] = E\left[\epsilon_t^2 \mathcal{K}_{\psi,t}(\pi, \lambda)\mathcal{K}_{\psi,t}(\tilde{\pi}, \tilde{\lambda})\right] = \sigma^2 E\left[\mathcal{K}_{\psi,t}(\pi, \lambda)\mathcal{K}_{\psi,t}(\tilde{\pi}, \tilde{\lambda})\right].$$

Further, $\sup_{\pi\in\Pi,\lambda\in\Lambda}|\mathfrak{X}_n(\pi, \lambda)| \xrightarrow{p} 0$ by Lemma B.13. This proves (B.11).

Stochastic equicontinuity for $\mathcal{E}_{\psi,t}(\psi_n, \chi)$ holds if $\forall(\epsilon, \eta) > 0$ there exists $\delta > 0$ such that:

$$\lim_{n\to\infty} \mathcal{P}_n(r, \delta, \eta) = \lim_{n\to\infty} P\left(\sup_{\chi,\tilde{\chi}\in\mathcal{X}:||\chi-\tilde{\chi}||\leq\delta} |\mathcal{E}_{\psi,t}(\psi_n, \chi) - \mathcal{E}_{\psi,t}(\psi_n, \tilde{\chi})| > \eta\right) < \epsilon. \tag{B.12}$$

We again adapt arguments in Arcones and Yu (1994, proof of Theorem 2.1 and Lemma 2.1) in order to verify (B.12). $\mathcal{E}_{\psi,t}(\psi_n, \chi)$ lies in the *V-C subgraph* class of functions $\mathcal{V}(\mathcal{C})$ because it is continuous, hence the covering numbers satisfy $\mathcal{N}(\varepsilon, \mathcal{K}, ||\cdot||_2) < a\varepsilon^{-b}$ for all $\varepsilon \in (0,1)$ and some $a, b > 0$ (e.g. Lemma 7.13 in Dudley, 1978, and Lemma II.25 in Pollard, 1984). Furthermore, under Assumption 1.b,c and by multiple uses of Minkowski and Hölder's inequalities, it is easily verified that $\mathcal{E}_{\psi,t}(\psi_n, \chi)$ is L_r-bounded, $r \equiv p/2 > 2$, and β-mixing with coefficients $\beta_l = O(l^{-qp/(q-p)-\iota})$, $q > p > 6$ and tiny $\iota > 0$. By simple algebra it follows $\beta_l = O(l^{-r/(r-2)}) = O(l^{-p/(p-4)})$ because $p/(p-4) < qp/(q-p)$. . Therefore $\{\mathcal{E}_{\psi,t}(\psi_n, \chi) : \chi \in \mathcal{X}\}$ is stochastically equicontinuous Arcones and Yu (1994, Lemma 2.1, see especially eq. (2.13)). $\mathcal{QED}$

Lemma B.10. *Under Assumption 1,* $\sup_{\omega \in \mathbb{R}^{k_\beta}:||\omega||=1, \pi\in\Pi, \lambda\in\Lambda} ||\hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda) - \mathfrak{b}_\theta(\omega,\pi,\lambda)|| \xrightarrow{p} 0$ *and* $\sup_{\pi\in\Pi,\lambda\in\Lambda} ||\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda) - \mathfrak{b}_\psi(\pi,\lambda)|| \xrightarrow{p} 0$.

Proof. Pointwise $\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda) \xrightarrow{p} \mathfrak{b}_\psi(\pi,\lambda)$ follows from stationarity, ergodicity, and the Assumption 1 moment bounds. It remains to show stochastic equicontinuity: $\forall(\epsilon,\eta) > 0$ there exists $\delta > 0$ such that:

$$\lim_{n\to\infty} \mathcal{P}_n(r,\delta,\eta) = \lim_{n\to\infty} P\left(\sup_{\chi,\tilde{\chi}\in\mathcal{X}:||\chi-\tilde{\chi}||\leq\delta} \left|\left\{\hat{\mathfrak{b}}_{\psi,n}(\chi) - \mathfrak{b}_\psi(\chi)\right\} - \left\{\hat{\mathfrak{b}}_{\psi,n}(\tilde{\chi}) - \mathfrak{b}_\psi(\tilde{\chi})\right\}\right| > \eta\right) < \epsilon.$$

where $\chi = [\lambda', \pi']' \in \mathcal{X} = \Lambda \times \Pi$. There exists $\chi_* \in \mathcal{X}$, $||\chi - \chi_*|| \leq ||\chi - \tilde{\chi}||$, such that:

$$\left\{\hat{\mathfrak{b}}_{\psi,n}(\chi) - \mathfrak{b}_\psi(\chi)\right\} - \left\{\hat{\mathfrak{b}}_{\psi,n}(\tilde{\chi}) - \mathfrak{b}_\psi(\tilde{\chi})\right\}$$
$$= \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\chi}\left\{(F(\lambda_*'\mathcal{W}(x_t))d_{\psi,t}(\pi_*) - E\left[F\left(\lambda_*'\mathcal{W}(x_t)\right)d_{\psi,t}(\pi_*)\right])\right\}'(\chi - \tilde{\chi}).$$

The envelop moment bounds in Assumption 1 imply:

$$E\left[\sup_{\chi\in\mathcal{X}}\left|\frac{\partial}{\partial\chi}\left\{(F(\lambda_*'\mathcal{W}(x_t))d_{\psi,t}(\pi_*) - E\left[F\left(\lambda_*'\mathcal{W}(x_t)\right)d_{\psi,t}(\pi_*)\right])\right\}\right|\right] \leq K < \infty.$$

Now invoke Markov's inequality to deduce $\mathcal{P}_n(r,\delta,\eta) \leq \eta^{-1}K\delta < \epsilon$ for any $0 < \delta < \epsilon\eta/K$. $\mathcal{QED}$

Define $\Theta^+ \equiv \{\theta^+ \in \mathbb{R}^{k_x+k_\beta+k_\pi+1} : \theta^+ = [||\beta||, \omega(\beta), \zeta, \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\}$ and

$$\epsilon_t(\theta^+) \equiv y_t - \zeta' x_t - \|\beta\|\,\omega' g(x_t,\pi) \quad\text{and}\quad \widehat{\mathcal{H}}_n(\omega,\pi) = \frac{1}{n}\sum_{t=1}^n d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)'$$
$$\hat{v}_n^2(\theta^+,\lambda) = \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta^+)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\theta,n}(\theta,\omega,\lambda)'\widehat{\mathcal{H}}_n^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2$$
$$v^2(\theta^+,\lambda) = E\left[\epsilon_t^2(\theta)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\theta,\omega,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2\right].$$

Lemma B.11. *Under Assumption 1,* $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda} ||\hat{v}_n^2(\theta^+,\lambda) - v^2(\theta^+,\lambda)|| \xrightarrow{p} 0$ *and* $\sup_{\theta\in\Theta,\lambda\in\Lambda} ||\hat{v}_n^2(\theta,\lambda) - v^2(\theta,\lambda)|| \xrightarrow{p} 0$.

Proof. We only prove $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda} ||\hat{v}_n^2(\theta^+,\lambda) - v^2(\theta^+,\lambda)|| \xrightarrow{p} 0$; the proof of $\sup_{\theta\in\Theta,\lambda\in\Lambda} ||\hat{v}_n^2(\theta,\lambda) - v^2(\theta,\lambda)|| \xrightarrow{p} 0$ is similar.

Define

$$v_n^2(\theta^+,\lambda) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta)\left\{F(\lambda'\mathcal{W}(x_t)) - \mathfrak{b}_\theta(\theta,\omega,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2$$

$$\mathcal{C}_n(\theta^+) \equiv \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)$$

$$\mathcal{E}_n(\theta^+,\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)F(\lambda\mathcal{W}(x_t))$$

Then:

$$\begin{aligned}\hat{v}_n^2(\theta^+,\lambda) - v_n^2(\theta^+,\lambda) \;&=\; -\left\{\hat{\mathfrak{b}}_{\theta,n}(\theta,\omega,\lambda)'\widehat{\mathcal{H}}_n^{-1}(\omega,\pi) - \mathfrak{b}_\theta(\theta,\omega,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)\right\} \\ &\quad\times\left\{2\mathcal{E}_n(\theta^+,\lambda) - \left(\hat{\mathfrak{b}}_{\theta,n}(\theta,\omega,\lambda)'\widehat{\mathcal{H}}_n^{-1}(\omega,\pi) + \mathfrak{b}_\theta(\theta,\omega,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)\right)\mathcal{C}_n(\theta^+)\right\}.\end{aligned}$$

By the same arguments used to prove Lemma B.6, $\mathcal{C}_n(\theta^+) \xrightarrow{p} E[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)]$ uniformly on Θ^+. Further, $\mathcal{E}_n(\theta^+,\lambda) \xrightarrow{p} E[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)F(\lambda'\mathcal{W}(x_t))]$ uniformly on $\Theta^+\times\Lambda$ because (i) pointwise convergence follows from the assumed moment and mixing properties, and (ii) $\mathcal{E}_n(\theta^+,\lambda)$ is stochastically equicontinuous by arguments in the proof of Lemma B.10 after simple alterations. Now apply Lemmas B.6 and B.10 to yield $|\hat{v}_n^2(\theta^+,\lambda) - v_n^2(\theta^+,\lambda)| \xrightarrow{p} 0$ uniformly on Θ^+. Finally, $v_n^2(\theta^+,\lambda) \xrightarrow{p} v^2(\theta^+,\lambda)$ uniformly on Θ^+ by the same arguments in the proof of Lemma B.10. $\mathcal{QED}$

Recall $\mathfrak{b}_\theta(\omega,\pi,\lambda) \equiv E[F(\lambda'\mathcal{W}(x_t))\,d_{\theta,t}(\omega,\pi)]$, and define

$$v^2(\lambda) \equiv v^2(\omega_0,\pi_0,\lambda)$$

where:

$$v^2(\omega,\pi,\lambda) \equiv E\left[\epsilon_t^2(\psi_0,\pi)\left\{F(\lambda'\mathcal{W}(x_t)) - \mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2\right].$$

Lemma B.12. *Let Assumptions 1.a(i) andAssumption 3 hold. Under* $\mathcal{C}(i,b)$ *with* $||b|| < \infty$*, the* *following set has Lebesgue measure zero:*

$$\left\{\lambda \in \Lambda : \inf_{\omega'\omega=1,\pi\in\Pi} v^2(\omega,\pi,\lambda) = 0\right\}.$$

Under $\mathcal{C}(ii,\omega_0)$, the set $\{\lambda \in \Lambda : v^2(\lambda) = 0\}$ has Lebesgue measure zero.

Proof. In view of $E[\epsilon_t^2|x_t] = \sigma_0^2 > 0$ $a.s.$ under Assumption 1.a(i), the proof under $\mathcal{C}(ii,\omega_0)$ is identical to Bieren's (1990, Lemma 2).

Consider weak identification cases $\mathcal{C}(i,b)$ with $||b|| < \infty$. Assume

$$S^* \equiv \left\{\lambda \in \Lambda : \inf_{\omega'\omega=1,\pi\in\Pi} v^2(\omega,\pi,\lambda) = 0\right\}$$

has positive Lebesgue measure, and take any $\lambda \in S^*$. Use $P(E[\inf_{\pi\in\Pi}\{\epsilon_t^2(\psi_0,\pi)\}|x_t] > 0) = 1$ under Assumption 3 to deduce

$$F(\lambda'\mathcal{W}(x_t)) = \mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\ a.s.$$

Now use the Assumption 3.b Borel function μ to yield that

$$E[\mu(x_t)F(\lambda'\mathcal{W}(x_t))] = E[\mu(x_t)d_{\theta,t}(\omega,\pi)']\mathcal{H}_\theta^{-1}(\omega,\pi)\mathfrak{b}_\theta(\omega,\pi,\lambda).$$

Note $\mathfrak{b}_\theta(\omega,\pi,\lambda) \equiv E[d_{\theta,t}(\omega,\pi)F(\lambda'\mathcal{W}(x_t))]$ hence

$$E[\mu(x_t)F(\lambda'\mathcal{W}(x_t))] = E[\xi(\omega,\pi)'d_{\theta,t}(\omega,\pi) \times F(\lambda'\mathcal{W}(x_t))],$$

where $\xi(\omega,\pi) \equiv \mathcal{H}_\theta^{-1}(\omega,\pi)E[\mu(x_t)d_{\theta,t}(\omega,\pi)]$. This implies

$$E[\{\mu(x_t) - \xi(\omega,\pi)'d_{\theta,t}(\omega,\pi)\}F(\lambda'\mathcal{W}(x_t))] = 0. \tag{B.13}$$

Since S^* has positive Lebesgue measure, the equality in (B.13) applies for all λ in a subset with positive Lebesgue measure. Thus $\mu(x_t) = \xi(\omega,\pi)'d_{\theta,t}(\omega,\pi)$ $a.s.$ by Theorem 2.3 in Stinchcombe and White (1998). Hence $E[\kappa_t(\omega,\pi)\kappa_t(\omega,\pi)']$ is singular, where $\kappa_t(\omega,\pi) \equiv [\mu(x_t), d_{\theta,t}(\omega,\pi)]'$, which contradicts Assumption 3.b(ii). $\mathcal{QED}$

Define

$$\mathcal{M}_t(\pi,\lambda) \equiv \{g(x_t,\pi_0) - g(x_t,\pi)\}F(\lambda'\mathcal{W}(x_t)) \text{ and } \tilde{\mathcal{M}}_t(\pi) \equiv \{g(x_t,\pi) - g(x_t,\pi_0)\}d_{\psi,t}(\pi)'.$$

Lemma B.13. *Under Assumption 1:*

$$\sup_{\pi\in\Pi,\lambda\in\Lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t F(\lambda'\mathcal{W}(x_t)) - E[\epsilon_t F(\lambda'\mathcal{W}(x_t))]\right| \xrightarrow{p} 0,$$
$$\sup_{\pi\in\Pi,\lambda\in\Lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\mathcal{M}_t(\pi,\lambda) - E[\mathcal{M}_t(\pi,\lambda)]\right| \xrightarrow{p} 0 \text{ where } \sup_{\pi\in\Pi,\lambda\in\Lambda}|E[\mathcal{M}_t(\pi,\lambda)]| < \infty$$
$$\sup_{\pi\in\Pi}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{\mathcal{M}}_t(\pi) - E[\tilde{\mathcal{M}}_t(\pi)]\right| \xrightarrow{p} 0 \text{ where } \sup_{\pi\in\Pi}\left|E[\tilde{\mathcal{M}}_t(\pi)]\right| < \infty.$$

Proof. In view of envelope moment bounds in Assumption 1.c, the argument is essentially identical to the proof of Lemma B.10. $\mathcal{QED}$.

C Proof of Theorem 4.1

Theorem 4.1. *Let Assumptions 1 and 2 hold.*

a. Under drift case $\mathcal{C}(i,b)$ with $||b|| < \infty$, $(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n), \hat{\pi}_n) \xrightarrow{d} (\tau(\pi^*(b),b), \pi^*(b))$.

b. Under drift case $\mathcal{C}(ii,\omega_0)$, $\sqrt{n}\mathfrak{B}(\hat{\beta}_n)(\hat{\theta}_n - \theta_n) \xrightarrow{d} -\mathcal{H}_\theta^{-1}\mathcal{G}_\theta$.

Proof.

Claim a.

Step 1: We first prove

$$\left\{\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n : \Pi\right\} \Rightarrow^* \{\tau(\pi,b) : \Pi\}. \tag{C.14}$$

Recall $\psi_{0,n} = [0'_{k_\beta}, \zeta'_0]'$. By Lemma B.7.a:

$$\begin{aligned}\sqrt{n}\left(\hat{\psi}_n(\pi) - \psi_n\right) &= \sqrt{n}\left(\hat{\psi}_n(\pi) - \psi_{0,n}\right) + \sqrt{n}\left(\psi_{0,n} - \psi_n\right) \\ &= -\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)\sqrt{n}\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi) - \left[\sqrt{n}\beta'_n, 0'_{k_\beta}\right]'.\end{aligned} \tag{C.15}$$

By the construction of $\mathcal{G}_{\psi,n}(\theta)$ in (A.3), we can write:

$$\sqrt{n}\frac{\partial}{\partial\psi}Q_n(\psi_{0,n},\pi) = \mathcal{G}_{\psi,n}(\psi_{0,n},\pi) - \sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]. \tag{C.16}$$

Assumption 1.a implies $E[\epsilon_t(\theta_n)d_{\psi,t}(\pi)] = 0$, hence:

$$\sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right] = \sqrt{n}E\left[\{\epsilon_t(\psi_{0,n},\pi) - \epsilon_t(\theta_n)\}d_{\psi,t}(\pi)\right] = E\left[\sqrt{n}\beta'_n g(x_t,\pi_0)d_{\psi,t}(\pi)\right]. \tag{C.17}$$

Therefore, by the definition of $\mathcal{D}_\psi(\pi)$ in (A.4), and $\sqrt{n}\beta_n \to b$ with $||b|| < \infty$:

$$\sup_{\pi\in\Pi}\left|\sqrt{n}E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right] + \mathcal{D}_\psi(\pi)b\right| \to 0. \tag{C.18}$$

By Lemma B.2 $\sup_{\pi\in\Pi}||\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)|| \xrightarrow{p} 0$, where $\mathcal{H}_\psi(\pi)$ is bounded and positive definite uniformly on Π. Now combine (C.15)-(C.18) to yield:

$$\sup_{\pi\in\Pi}\left\|\sqrt{n}\left(\hat{\psi}_n(\pi) - \psi_n\right) - \left(-\mathcal{H}_\psi^{-1}(\pi)\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\} - \left[b, 0'_{k_\beta}\right]'\right)\right\| \xrightarrow{p} 0. \tag{C.19}$$

Therefore (C.14) follows by application of Lemma B.1.

Step 2: Now turn to $\hat{\pi}_n$. Write $Q_n^c(\pi) \equiv Q_n(\hat{\psi}_n(\pi), \pi)$. Let drift case $\mathcal{C}(i, b)$ hold with $||b|| < \infty$. By Lemma B.8.a $\{n(Q_n^c(\pi), \pi) - Q_{0,n}) : \Pi\} \Rightarrow^* \{\xi_\psi(\pi, b) : \Pi\}$, hence by the mapping theorem $|\arg\min_{\pi\in\Pi}\{n(Q_n^c(\pi) - Q_{0,n})\} - \arg\min_{\pi\in\Pi}\{\xi_\psi(\pi, b)\}| \xrightarrow{p} 0$. Therefore $\hat{\pi}_n \xrightarrow{d} \pi^*(b) = \arg\min_{\pi\in\Pi}\{\xi_\psi(\pi, b)\}$ by the mapping theorem and Assumption 2.

Step 3: The proof is complete by showing joint weak convergence for $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$ and $\hat{\pi}_n$.

First, $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$ and $\hat{\pi}_n$ are continuous functions of $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$. The former follows from (C.15) and (C.16). In order to understand $\hat{\pi}_n$, define

$$\xi_{\psi,n}(\pi, b) \equiv -\frac{1}{2}\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\}' \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)\left\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\}.$$

By Lemmas B.1 and B.2 $\{\xi_{\psi,n}(\pi, b) : \Pi\} \Rightarrow^* \{\xi_\psi(\pi, b) : \Pi\}$. Hence, by Lemma B.8.a and the mapping theorem

$$\left|\arg\min_{\pi\in\Pi}\left\{n(Q_n^c(\pi) - Q_{0,n})\right\} - \arg\min_{\pi\in\Pi}\left\{\xi_{\psi,n}(\pi, b)\right\}\right| \xrightarrow{p} 0.$$

In view of the argument above, this implies

$$\left|\hat{\pi}_n - \arg\min_{\pi\in\Pi}\left\{\xi_{\psi,n}(\pi, b)\right\}\right| \xrightarrow{p} 0.$$

Hence $\hat{\pi}_n$ can be expressed as a continuous function of $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$.

Second, $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ converge jointly because the latter has a non-random limit uniformly on Π (cf. Andrews and Cheng, 2012b, p. 25). Hence

$$\left\{\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n : \Pi\right\} \Rightarrow^* \left\{\tau(\pi, b), \pi^*(b) : \Pi\right\}.$$

By the mapping theorem it therefore follows that:

$$\left(\sqrt{n}\hat{\psi}_n(\hat{\pi}_n) - \psi_n), \hat{\pi}_n\right) \Rightarrow^* (\tau(\pi^*(b), b), \pi^*(b)) .$$

Finally, a subsequent proof requires uniform consistency

$$\sup_{\pi \in \Pi} \left\| \hat{\psi}_n(\pi) - \psi_n \right\| \xrightarrow{p} 0. \tag{C.20}$$

Note that

$$\hat{\psi}_n(\pi) - \psi_n = -\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)\frac{\partial}{\partial \psi}Q_n(\psi_{0,n}, \pi) - \left[\beta_n', 0_{k_\beta}'\right]',$$

where $\sup_{\pi \in \Pi} ||\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi) - \mathcal{H}^{-1}_{\psi}(\pi)|| \xrightarrow{p} 0$ and $\beta_n \to 0$. Moreover, by the Assumption 1.b,c,d(iii) moment and envelope bounds and $\beta_n \to 0$:

$$\begin{aligned}
\sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) \right\| &\leq \sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - E\left[\epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi)\right] \right\| + \sup_{\pi \in \Pi} \left\| E\left[\beta_n' g(x_t, \pi_0) d_{\psi,t}(\pi)\right] \right\| \\
&= \sup_{\pi \in \Pi} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi) - E\left[\epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi)\right] \right\| + o_p(1) \\
&\equiv \mathfrak{E}_n + o_p(1).
\end{aligned}$$

Finally, $\mathfrak{E}_n \xrightarrow{p} 0$ by the same arguments used to prove Lemmas B.2 and B.6. Therefore:

$$\begin{aligned}
\sup_{\pi \in \Pi} \left\| \hat{\psi}_n(\pi) - \psi_n \right\| &= \sup_{\pi \in \Pi} \left\| -\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi) \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - \left[\beta_n', 0_{k_\beta}'\right] \right\| \\
&\leq \sup_{\pi \in \Pi} \left\| -\mathcal{H}^{-1}_{\psi}(\pi) \mathfrak{E}_n - \left[\beta_n', 0_{k_\beta}'\right] \right\| + o_p(1) \xrightarrow{p} 0.
\end{aligned}$$

This proves (C.20).

Claim b. Let drift case $\mathcal{C}(ii, \omega_0)$ hold, and define

$$\widehat{\mathcal{H}}_n(\omega, \pi) \equiv \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \text{ and } \mathcal{H}_\theta(\omega, \pi) \equiv E\left[d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)'\right].$$

Recall $\mathfrak{B}(\beta)$ defined in (A.2) and $\omega(\beta)$ defined in (A.5). By the first order condition $(\partial/\partial\theta)Q_n(\hat{\theta}_n) = 0$ and the mean value theorem there exists θ_n^*, $||\theta_n^* - \theta_n|| \leq ||\hat{\theta}_n - \theta_n||$, such that:

$$\begin{aligned}
0 &= \mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) + \mathfrak{B}(\beta_n)^{-1} \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) \mathfrak{B}(\beta_n)^{-1} \times \sqrt{n} \mathfrak{B}(\beta_n) \left(\hat{\theta}_n - \theta_n\right) \\
&= \mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) + \widehat{\mathcal{H}}_n(\omega(\beta_n^*), \pi_n^*) \sqrt{n} \mathfrak{B}(\beta_n) \left(\hat{\theta}_n - \theta_n\right).
\end{aligned}$$

The second equality follows from the constructions of $\mathfrak{B}(\beta)$, $(\partial^2/\partial\theta\partial\theta')Q_n(\theta)$ and $\widehat{\mathcal{H}}_n(\theta)$. Hence:

$$\sqrt{n}\mathfrak{B}(\beta_n)\left(\hat{\theta}_n - \theta_n\right) = \widehat{\mathcal{H}}_n^{-1}(\omega(\beta_n^*), \pi_n^*)\mathfrak{B}(\beta_n)^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n). \tag{C.21}$$

Observe that $||\theta_n^* - \theta_n|| \leq ||\hat{\theta}_n - \theta_n||$, and by the argument below:

$$\left\|\hat{\theta}_n - \theta_n\right\| \xrightarrow{p} 0. \tag{C.22}$$

Hence $\widehat{\mathcal{H}}_n(\omega(\beta_n^*), \pi_n^*) \xrightarrow{p} \mathcal{H}_\theta$ by Lemma B.6 and continuity. Corollary B.4 now yields the result.

It remains to prove (C.22). Use (C.21) to yield:

$$\begin{aligned}\left\|\sqrt{n}\mathfrak{B}(\beta_n)\left(\hat{\theta}_n - \theta_n\right)\right\| \leq & \sup_{\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi}\left\|\widehat{\mathcal{H}}_n^{-1}(\omega,\pi) - \mathcal{H}_\theta^{-1}(\omega,\pi)\right\|\left\|\mathfrak{B}(\beta_n)^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n)\right\| \\ & + \sup_{\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi}\left\|\mathcal{H}_\theta^{-1}(\omega,\pi)\right\|\left\|\mathfrak{B}(\beta_n)^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n)\right\|.\end{aligned}$$

By Lemma B.6 and the Slutsky Theorem

$$\sup_{\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi}\left\|\widehat{\mathcal{H}}_n^{-1}(\omega,\pi) - \mathcal{H}_\theta^{-1}(\omega,\pi)\right\| \xrightarrow{p} 0,$$

where $\sup_{\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi}||\mathcal{H}_\theta^{-1}(\omega,\pi)|| < \infty$ follows from the eigenvalue bounds in Lemma B.6. Moreover, by Lemma B.3 and the mapping theorem

$$\mathfrak{B}(\beta_n)^{-1}\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_n) = O_p(1).$$

This proves $\sqrt{n}\mathfrak{B}(\beta_n)(\hat{\theta}_n - \theta_n) = O_p(1)$, hence (C.22). $\mathcal{QED}$.

D Identification Category Selection Type 2 P-Value

Operate under H_0. Define $\mathcal{F}_\infty(c) \equiv P(\mathcal{T}(\lambda) \leq c)$ where $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$ is the asymptotic null chi-squared process under strong identification, and let $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda, h) \leq c)$ where $\{\mathcal{T}_\psi(\lambda, h) : \lambda \in \Lambda\}$ is the asymptotic null process under weak identification. The case specific asymptotic p-values are

$$p_n^\infty(\lambda) \equiv 1 - \mathcal{F}_\infty(\mathcal{T}_n(\lambda)) = \bar{\mathcal{F}}_\infty(\mathcal{T}_n(\lambda)) \text{ and } p_n(\lambda, h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda)) = \bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda)).$$

The ICS-2 p-value is computed as follows. Let $(\Delta_1, \Delta_2) \in [0, 1)$ and $\kappa > 0$ be user chosen numbers. Let s be a continuous function on $[0, \infty)$, such that $s(x) \in [0,1]$, $s(x)$ is non-increasing in x, $s(0) = 1$, and $s(x) \to 0$ as $x \to \infty$. Then, using $\mathcal{A}_n$ in (A.6):

$$p_n^{(ICS-2)}(\lambda) = \left\{ \; p_{n,1}(\lambda; \Delta_1) \text{ if } \mathcal{A}_n \leq \kappa, \quad p_{n,2}(\lambda, \Delta_1, \Delta_2) \text{ if } \mathcal{A}_n > \kappa \right.$$

where

$$p_{n,1}(\lambda; \Delta_1) \equiv \max \left\{ \sup_{h \in \mathfrak{H}} \{p_n(\lambda, h)\}, p_n^{\infty}(\lambda) \right\} + \Delta_1 \tag{D.23}$$
$$p_{n,2}(\lambda, \Delta_1, \Delta_2) \equiv p_n^{\infty}(\lambda) + \Delta_2 + \{p_{n,1}(\lambda; \Delta_1) - p_n^{\infty}(\lambda) - \Delta_2\} \, s(\mathcal{A}_n - \kappa).$$

The construction allows for a smooth transition between identification cases, and allows for a non-diverging threshold. The latter necessitates the tuning parameters (Δ_1, Δ_2) which promote a correct asymptotic size. See also Andrews and Barwick (2012) for a related method.

See Andrews and Cheng (2012a, p. 2193) for details on determining appropriate choices for $(\Delta_1, \Delta_2, \kappa)$. In theory $\kappa > 0$ can be any value since the ICS-2 p-value $p_n^{(ICS-2)}(\lambda)$ promotes a test with correct asymptotic level. Andrews and Cheng (2012a, p. 2194) and Andrews and Cheng (2013a, p. 50) choose κ for robust t-statistics by minimizing the False Coverage Probability [FCP] for the corresponding robust confidence set.[2] The CM test statistic is not based on a parametric hypothesis, hence the FCP method does not apply. Instead, we may choose ad hoc values like $\kappa = 1$ or $\kappa = 1.5$, based on finite sample experiments for various models.[3] Since our focus is an asymptotically valid method for computing $p_n(\lambda, h)$, and therefore $\{p_n^{(LF)}(\lambda), p_n^{(ICS-2)}(\lambda), p_n^{(ICS-2)}(\lambda)\}$, we do not present here a theory based alternative to minimizing the FCP in order to select κ for CM tests.

We choose (Δ_1, Δ_2) to ensure the asymptotic Null Rejection Probability [NRP] under weak identification $\sqrt{n}||\beta_n|| \to [0, \infty)$ is not larger than α (Andrews and Cheng, 2012a, Section 5.3). The NRP is

$$NRP_n(\Delta_1, \Delta_2; \lambda,) \equiv P\left(p_{n,1}(\lambda; \Delta_1) \leq \alpha \cap \mathcal{A}_n \leq \kappa\right) + P\left(p_{n,2}(\lambda; \Delta_1, \Delta_2) \leq \alpha \cap \mathcal{A}_n > \kappa\right).$$

Note that $\mathcal{A}_n \xrightarrow{d} \mathcal{A}(b)$ under weak identification, where $\mathcal{A}(b)$ is defined in Theorem 5.1.a. Under strong identification and regularity conditions, $\mathcal{A}_n \xrightarrow{p} \infty$ (Theorem 5.1.b).

[2] Consider the parametric hypothesis $\mathcal{R}(\theta) = 0$. The FCP of a confidence set for $\mathcal{R}(\theta)$ is the probability that the confidence set contains a value different from the true $\mathcal{R}(\theta_n)$, where $\theta_n \equiv [\beta_n', \zeta_0', \pi_0']'$.

[3] Andrews and Cheng (2012a,b, 2013a,b) find that a wide range of values for κ lead to similar results for robust Smooth Transition Autoregression model based t-tests, including $\kappa = 1$ and $\kappa = 1.5$, because Δ_1 and Δ_2 are computed to ensure correct asymptotic size for any chosen κ.

Define

$$p_1(\lambda,\tilde{h};\Delta_1) \equiv \max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\},\bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\}+\Delta_1 \tag{D.24}$$
$$p_2(\lambda,\tilde{h};\Delta_1,\Delta_2) \equiv \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))+\Delta_2+\left\{p_1(\lambda)-\bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))-\Delta_2\right\}s\left(\mathcal{A}(b)-\kappa\right).$$

$\sup_{h\in\mathfrak{H}}$ operates on the distribution function $\bar{\mathcal{F}}_{\lambda,h}$ and not its argument $\mathcal{T}_\psi(\lambda,\tilde{h})$. This follows from the definition $p_{n,1}(\lambda;\Delta_1)\equiv\max\{\sup_{h\in\mathfrak{H}}\{p_n(\lambda,h)\},p_n^\infty(\lambda)\}+\Delta_1$, and under weak identification:

$$\sup_{h\in\mathfrak{H}}\{p_n(\lambda,h):\lambda\in\Lambda\}=\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda)):\lambda\in\Lambda\right\}\Rightarrow^*\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h})):\lambda\in\Lambda\right\}.$$

By Theorem 6.1 and the mapping theorem:

$$\{p_{n,1}(\lambda;\Delta_1):\lambda\in\Lambda\}\Rightarrow^*\left\{p_1(\lambda,\tilde{h};\Delta_1):\lambda\in\Lambda\right\}$$

and

$$\{p_{n,2}(\lambda;\Delta_1,\Delta_2):\lambda\in\Lambda\}\Rightarrow^*\left\{p_2(\lambda,\tilde{h};\Delta_1,\Delta_2):\lambda\in\Lambda\right\}.$$

Joint convergence for $(p_{n,1}(\lambda;\Delta_1),p_{n,2}(\lambda;\Delta_1,\Delta_2),\mathcal{A}_n)$ is straightforward to prove: see the proof of Theorem 6.2. The asymptotic NRP under weak identification is therefore:

$$\begin{aligned} NRP(\Delta_1,\Delta_2;\lambda,\tilde{h}) \;\equiv\;& P\left(p_1(\lambda,\tilde{h};\Delta_1)<\alpha\cap\mathcal{A}(b)\leq\kappa\right) \\ &+P\left(p_2(\lambda,\tilde{h};\Delta_1,\Delta_2)<\alpha\cap\mathcal{A}(b)>\kappa\right). \end{aligned} \tag{D.25}$$

The role (Δ_1,Δ_2) play are the same as in Andrews and Cheng (2012a, p. 2193). Let $\tilde{b}_{\sup}$ be such that

$$\tilde{h}_{\sup}\equiv\left[\tilde{b}_{\sup},\tilde{\gamma}_{\sup}\right]=\arg\sup_{\tilde{h}\in\mathfrak{H}}\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}\left(\mathcal{T}_\psi(\lambda,\tilde{h})\right)\right\},$$

and $C\geq 0$ is some constant, e.g. $C=1$. Define the set

$$\mathfrak{H}_1\equiv\left\{h=[b,\gamma]:h\in\mathfrak{H},\;\;\|b\|\leq\left\|\tilde{b}_{\sup}\right\|+C\right\},$$

and define

$$\Delta_1\;\equiv\;\sup_{\tilde{h}\in\mathfrak{H}_1}\Delta_1(\tilde{h})\text{ where }\left\{\begin{array}{l}\Delta_1(\tilde{h})\geq 0\text{ solves }NRP(\Delta_1(\tilde{h}),0;\tilde{h})=\alpha\\ \Delta_1(\tilde{h})=0\text{ if }NRP(0,0;\tilde{h})<\alpha\end{array}\right.$$

$$\Delta_2 \equiv \sup_{\tilde{h}\in\mathfrak{H}_1} \Delta_2(\tilde{h}) \text{ where } \begin{cases} \Delta_2(\tilde{h}) \geq 0 \text{ solves } NRP(\Delta_1, \Delta_2(\tilde{h}); \tilde{h}) = \alpha \\ \Delta_1(\tilde{h}) = 0 \text{ if } NRP(\Delta_1, 0; \tilde{h}) < \alpha \end{cases}.$$

If $NRP(\Delta_1, 0; \tilde{h}) = \alpha$ does not hold for any Δ_1, then choose any Δ_1 that satisfies $NRP(\Delta_1, 0; \tilde{h}) \leq \alpha$. The following lemma shows the latter is always feasible (see the proof for examples). Thus, $NRP(\Delta_1, 0; \tilde{h}) = \alpha$ for some Δ_1 holds when $NRP(\Delta_1, 0; \tilde{h})$ is strictly decreasing and continuous in Δ_1, which generally holds in view of the construction of $\mathcal{T}_\psi(\lambda, \tilde{h})$. Similar derivations apply to Δ_2.

Lemma D.1. *Let $\sqrt{n}||\beta_n|| \to [0, \infty)$, and assume $\mathcal{F}_{\lambda,h}(c)$ is continuous* a.e. *on $[0, \infty)$. There always exists a (possibly non-unique) Δ_1 such that $\sup_{\tilde{h}\in\mathfrak{H}} NRP(\Delta_1, 0; \tilde{h}) \leq \alpha$.*

Define

$$AsySz(\lambda) = \lim_{n\to\infty}\sup \sup_{\gamma\in\Gamma^*} P_\gamma \left(p_n^{(\cdot)}(\lambda) < \alpha | H_0\right).$$

Theorem D.2. *Let Assumptions 1-2, 4 and 5 hold. The ICS-2 $p_n^{(ICS-2)}(\lambda)$ satisfies $AsySz(\lambda) \leq \alpha$.*

Proof of Lemma D.1. By (D.25), the asymptotic Null Rejection Probability under $\sqrt{n}||\beta_n|| \to [0, \infty)$ is

$$NRP(\Delta_1, \Delta_2; \tilde{h}) = P\left(p_1(\lambda, \tilde{h}; \Delta_1) < \alpha \cap \mathcal{A}(b) \leq \kappa\right) + P\left(p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa\right). \quad \text{(D.26)}$$

Define $p^{(LF)}(\lambda, \tilde{h}) \equiv \max\{\sup_{h\in\mathfrak{H}}\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h}))\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h}))\}$. Note that

$$\begin{aligned} P\left(p_1(\lambda, \tilde{h}; \Delta_1) < \alpha \cap \mathcal{A}(b) \leq \kappa\right) &\leq P\left(p^{(LF)}(\lambda, \tilde{h}) < \alpha \cap \mathcal{A}(b) \leq \kappa\right) \\ &\leq P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h}))\right\} < \alpha \cap \mathcal{A}(b) \leq \kappa\right) \end{aligned} \quad \text{(D.27)}$$

and

$$\begin{aligned} &P\left(p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &= P\left(\bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) + \Delta_2 + \left\{p^{(LF)}(\lambda, \tilde{h}) + \Delta_1 - \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) - \Delta_2\right\} s\left(\mathcal{A}(b) - \kappa\right) < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &\leq P\left(\bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h}))\left(1 - s\left(\mathcal{A}(b) - \kappa\right)\right) + p^{(LF)}(\lambda, \tilde{h}) s\left(\mathcal{A}(b) - \kappa\right) + \Delta_1 s\left(\mathcal{A}(b) - \kappa\right) < \alpha \cap \mathcal{A}(b) > \kappa\right). \end{aligned}$$

Consider two examples:

$$\Delta_1(\tilde{h}) = \left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h}))\right\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h}))\right\} - \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h}))\right) \frac{1 - s\left(\mathcal{A}(b) - \kappa\right)}{s\left(\mathcal{A}(b) - \kappa\right)} \quad \text{(D.28)}$$

$$\Delta_1(\tilde{h}) = \max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\}\frac{(1-s(\mathcal{A}(b)-\kappa))}{s(\mathcal{A}(b)-\kappa)}. \tag{D.29}$$

Use $\Delta_1(\tilde{h})$ in (D.28) to yield

$$\begin{aligned} P\left(p_2(\lambda,\tilde{h};\Delta_1,\Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa\right) &\leq P\left(p^{(LF)}(\lambda,\tilde{h}) < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &\leq P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\} < \alpha \cap \mathcal{A}(b) > \kappa\right), \end{aligned}$$

hence

$$\begin{aligned} \sup_{\tilde{h}\in\mathfrak{H}} NRP(\Delta_1(\tilde{h}),0;\tilde{h}) &\leq \sup_{\tilde{h}\in\mathfrak{H}} P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\} < \alpha \cap \mathcal{A}(b) \leq \kappa\right) \\ &\quad + \sup_{\tilde{h}\in\mathfrak{H}} P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\} < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &= \sup_{\tilde{h}\in\mathfrak{H}} P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\} < \alpha\right) \leq \sup_{\tilde{h}\in\mathfrak{H}} P\left(\bar{\mathcal{F}}_{\lambda,\tilde{h}}(\mathcal{T}_\psi(\lambda,\tilde{h})) < \alpha\right) = \alpha. \end{aligned}$$

The final equality holds because $\bar{\mathcal{F}}_{\lambda,\tilde{h}}$ is continuous by assumption, and $\mathcal{T}_\psi(\lambda,\tilde{h})$ is distributed $\mathcal{F}_{\lambda,\tilde{h}}$.

Finally, note that

$$\begin{aligned} &P\left(p_2(\lambda,\tilde{h};\Delta_1,\Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &\quad\leq P\Big(\bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))\,(1-s(\mathcal{A}(b)-\kappa)) \\ &\qquad\quad + p^{(LF)}(\lambda,\tilde{h})s(\mathcal{A}(b)-\kappa) + \Delta_1 s(\mathcal{A}(b)-\kappa) < \alpha \cap \mathcal{A}(b) > \kappa\Big) \\ &\quad\leq P\left(p^{(LF)}(\lambda,\tilde{h})s(\mathcal{A}(b)-\kappa) + \Delta_1 s(\mathcal{A}(b)-\kappa) < \alpha \cap \mathcal{A}(b) > \kappa\right). \end{aligned}$$

Then using $\Delta_1(\tilde{h})$ in (D.29):

$$\begin{aligned} P\left(p_2(\lambda,\tilde{h};\Delta_1,\Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa\right) &\leq P\left(p^{(LF)}(\lambda,\tilde{h}) < \alpha \cap \mathcal{A}(b) > \kappa\right) \\ &\leq P\left(\bar{\mathcal{F}}_{\lambda,\tilde{h}}(\mathcal{T}_\psi(\lambda,\tilde{h})) < \alpha \cap \mathcal{A}(b) > \kappa\right). \end{aligned} \tag{D.30}$$

Combine (D.26), (D.27) and (D.30) to yield:

$$\sup_{\tilde{h}\in\mathfrak{H}} NRP(\Delta_1(\tilde{h}),0;\tilde{h}) \leq \sup_{\tilde{h}\in\mathfrak{H}} P\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\} < \alpha\right) = \alpha.$$

This completes the proof. $\mathcal{QED}$.

Proof of Theorem D.2.

Step 1. Under $\mathcal{C}(i,b)$ with $||b|| < \infty$, $\mathcal{A}_n \stackrel{d}{\to} \mathcal{A}(b)$ where $\mathcal{A}(b)$ is defined in Theorem 5.1.a. In Step 2 we show joint weak convergence under $\mathcal{C}(i,b)$

$$\{\mathcal{T}_n(\lambda), \mathcal{A}_n : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda, h), \mathcal{A}(b) : \Lambda\}. \tag{D.31}$$

Therefore, by the mapping theorem and Assumption 5:

$$\{p_{n,1}(\lambda; \Delta_1), p_{n,2}(\lambda; \Delta_1, \Delta_2), \mathcal{A}_n : \Lambda\} \Rightarrow^* \left\{p_1(\lambda, \tilde{h}; \Delta_1), p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2), \mathcal{A}(b) : \Lambda\right\}$$

where

$$p_1(\lambda, \tilde{h}; \Delta_1) = \max\left\{\sup_{h \in \mathfrak{H}} \left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h}))\right\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h}))\right\} + \Delta_1 \equiv p^{(LF)}(\lambda, \tilde{h}) + \Delta_1$$
$$p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) \equiv \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) + \Delta_2 + \left\{p_1 - \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) - \Delta_2\right\} s\left(\mathcal{A}(b) - \kappa\right).$$

The asymptotic size $AsySz(\lambda)$ is therefore

$$\begin{aligned}
&\lim_{n\to\infty}\sup \sup_{\gamma \in \Gamma^*} P_\gamma\left(p_n^{(ICS-2)}(\lambda) < \alpha | H_0\right) \\
&\quad = \sup_{\tilde{h} \in \mathfrak{H}} P\left(p_1(\lambda, \tilde{h}; \Delta_1) < \alpha \cap \mathcal{A}(b) \leq \kappa\right) + \sup_{\tilde{h} \in \mathfrak{H}} P\left(p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa | H_0\right) \\
&\quad = \sup_{\tilde{h} \in \mathfrak{H}} NRP(\Delta_1, \Delta_2; \lambda, \tilde{h}),
\end{aligned}$$

where NRP is the asymptotic Null Rejection Probability defined in (D.25). The tuning parameters (Δ_1, Δ_2) are chosen by supposition to ensure $\sup_{\tilde{h} \in \mathfrak{H}} NRP(\Delta_1, \Delta_2; \lambda, \tilde{h}) \leq \alpha$, cf. Lemma D.1.

Under $\mathcal{C}(ii, \omega_0)$ we have $\mathcal{A}_n \stackrel{p}{\to} \infty$ by Theorem 5.1.b. Hence $s(\mathcal{A}_n - \kappa) \stackrel{p}{\to} 0$ since the continuous function $s(x) \to 0$ as $x \to \infty$. Now apply Theorem 4.2.b and the mapping theorem to yield $\{p_{n,2}(\lambda; \Delta_1, \Delta_2) : \Lambda\} \Rightarrow^* \{\bar{\mathcal{F}}_\infty(\mathcal{T}(\lambda)) + \Delta_2 : \Lambda\}$. Since $\mathcal{T}(\lambda)$ is distributed $\mathcal{F}_\infty$, it therefore follows:

$$\begin{aligned}
AsySz(\lambda) &= \lim_{n\to\infty}\sup \sup_{\gamma \in \Gamma^*} P_\gamma\left(p_n^{(ICS-2)} < \alpha | H_0\right) \\
&= P\left(\bar{\mathcal{F}}_\infty(\mathcal{T}(\lambda)) + \Delta_2 < \alpha | H_0\right) \leq P\left(\bar{\mathcal{F}}_\infty(\mathcal{T}(\lambda)) < \alpha | H_0\right) = \alpha.
\end{aligned}$$

Step 2 (joint convergence). It remains to prove (D.31). Recall $\mathcal{S}_\beta \equiv [I_{k_\beta} : 0_{k_x \times k_x}]$, and define:

$$\omega(\hat{\beta}_n(\hat{\pi}_n)) = \frac{\sqrt{n}\mathcal{S}_\beta \hat{\psi}_n(\hat{\pi}_n)}{\left\|\sqrt{n}\mathcal{S}_\beta \hat{\psi}_n(\hat{\pi}_n)\right\|} = \frac{\sqrt{n}\mathcal{S}_\beta\left(\hat{\psi}_n(\hat{\pi}_n) - \psi_n\right) + \sqrt{n}\beta_n}{\left\|\sqrt{n}\mathcal{S}_\beta\left(\hat{\psi}_n(\hat{\pi}_n) - \psi_n\right) + \sqrt{n}\beta_n\right\|} \equiv \omega_n(\hat{\pi}_n),$$

hence $\omega_n(\hat{\pi}_n)$ is a continuous function of $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)$ and $\hat{\pi}_n$. By the argument leading to (A.12) in the proof of Theorem 4.2 in the main paper:

$$\sup_{\lambda \in \Lambda}\left|\mathcal{T}_n(\lambda) - \frac{(\mathfrak{Z}_n(\hat{\pi}_n, \lambda) + \mathcal{R}(\hat{\pi}_n, \lambda))^2}{v^2(\omega_n(\hat{\pi}_n), \hat{\pi}_n, \lambda)}\right| \xrightarrow{p} 0.$$

Recall $\{\mathcal{T}_n(\lambda) : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda, h) : \Lambda\}$ by Theorem 4.2.

By the proof of Theorem 5.1.a and the mapping theorem, $||\hat{\Sigma}_n - \bar{\Sigma}(\pi^*(b), b)|| \xrightarrow{p} 0$, where

$$\bar{\Sigma}(\pi, b) \equiv \Sigma(\omega^*(\pi, b), \pi) = \Sigma\left(\|\beta_0\|, \omega^*(\pi, b), \zeta_0, \pi\right),$$

and

$$\Sigma\left(\|\beta\|, \omega, \zeta, \pi\right) = \Sigma(\theta^+) \equiv \mathcal{H}_\theta(\theta^+)^{-1}\mathcal{V}(\theta^+)\mathcal{H}_\theta(\theta^+)^{-1}.$$

Therefore

$$\begin{aligned}\mathcal{A}_n &= \left(\frac{1}{p+1} n\hat{\beta}_n'\hat{\Sigma}_{\beta,\beta,n}^{-1}\hat{\beta}_n'\right)^{1/2} \\ &= \left(\frac{1}{p+1}\left(\mathcal{S}_\beta\sqrt{n}\left(\hat{\psi}_n - \psi_n\right) + \sqrt{n}\beta_n\right)'\bar{\Sigma}_{\beta,\beta}^{-1}(\hat{\pi}_n, b)\left(\mathcal{S}_\beta\sqrt{n}\left(\hat{\psi}_n - \psi_n\right) + \sqrt{n}\beta_n\right)\right)^{1/2} + o_p(1),\end{aligned}$$

where $\bar{\Sigma}_{\beta,\beta}(\pi, b)$ is the upper $(p+1) \times (p+1)$ block of $\bar{\Sigma}(\pi, b)$. Further $\mathcal{A}_n \xrightarrow{d} \mathcal{A}(b)$ by Theorem 5.1.a.

Therefore $\{\mathcal{T}_n(\lambda), \mathcal{A}_n : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda, h), \mathcal{A}(b) : \Lambda\}$ if we prove joint weak convergence for $(\mathfrak{Z}_n(\pi, \lambda), \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n)$ on $\Pi \times \Lambda$. By the proof of Theorem 4.1.a, $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$ and $\hat{\pi}_n$ are continuous functions of $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$, and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ has a constant limit in probability uniformly on Π by Lemma B.2. Joint weak convergence for $(\mathfrak{Z}_n(\pi, \lambda), \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n)$ therefore follows from joint weak convergence for $(\mathfrak{Z}_n(\pi, \lambda), \mathcal{G}_{\psi,n}(\psi_{0,n}, \pi))$, which is shown in Step 3 in the proof of Theorem 4.2. $\mathcal{QED}$.

E Robust Critical Values

We present bootstrapped identification category robust critical values. The idea is based on (unobserved) robust Least Favorable, and type 1 and 2 identification category selection [ICS] critical values presented in Andrews and Cheng (2012a).

E.1 Least Favorable and Identification Category Selection Critical Values

Let $\{\mathcal{T}_\psi(\lambda, b) : \lambda \in \Lambda\}$ denote the null limit process of $\mathcal{T}_n(\lambda)$ under weak identification $\sqrt{n}\beta_n \to b$ with $||b|| < \infty$ (see Theorem 5.2). Recall that ϕ_0 indexes all remaining (nuisance) parameters such that the distribution of $W_t \equiv [y_t, y_{t-1}, ..., y_{t-p}]'$ is determined by:

$$\gamma_0 \equiv (\theta_0, \phi_0) \in \Gamma^* \equiv \{\theta \in \Theta^*, \phi \in \Phi^*(\theta)\}. \tag{E.32}$$

Assume $\Phi^*(\theta) \subset \Phi^*\ \forall \theta \in \Theta^*$, where Φ^* is a compact metric space with some metric that induces weak convergence of the bivariate distributions of (W_t, W_{t+h}) for all t and $h \geq 1$.

Define the parametric set that characterizes data generating processes under weak identification $\beta_n \to \beta_0 = 0$, and $\sqrt{n}\beta_n \to b$ with $||b|| < \infty$:

$$h \equiv (\gamma_0, b) \in \mathfrak{H} \equiv \{h : \gamma_0 \in \Gamma^*, \text{ and } \|b\| < \infty, \text{ with } \beta_0 = 0\}. \tag{E.33}$$

Now let $\{\mathcal{T}_\psi(\lambda, h) : \lambda \in \Lambda\}$ denote the non-standard null limit process under weak identification. Under strong identification the null limit law is $\chi^2(1)$. Let $c_{1-\alpha}(\lambda, h)$ and $\chi^2_{1-\alpha}$ respectively be the $1 - \alpha$ quantiles for $\mathcal{T}_\psi(\lambda, h)$ and $\chi^2(1)$. All subsequent critical values are functions of $c_{1-\alpha}(\lambda, h)$, hence in Appendix E.3 we discuss how to compute $c_{1-\alpha}(\lambda, h)$ by bootstrap.

The following summarizes ideas developed in Andrews and Cheng (2012a, Section 5).

E.1.1 Least Favorable Critical Value

The *Least favorable* [LF] critical value is

$$c^{(LF)}_{1-\alpha}(\lambda) \equiv \max\left\{\sup_{h\in\mathfrak{H}}\{c_{1-\alpha}(\lambda, h)\}, \chi^2_{1-\alpha}\right\}.$$

A better critical value in terms of power uses the fact that (ζ_0, β_n) are consistently estimated by $(\hat{\zeta}_n, \hat{\beta}_n)$ under any degree of (non)identification. The *plug-in* LF critical value $\hat{c}^{(LF)}_{1-\alpha}(\lambda)$ uses $\widehat{\mathfrak{H}} \equiv \{h \in \mathfrak{H} : \theta = [\hat{\zeta}'_n, \hat{\beta}'_n, \pi']'\}$ in place of $\mathfrak{H}$.

In the present environment the null hypothesis is tested by using a sample version of $E[\epsilon_t F(\lambda' \mathcal{W}(x_t))]$. Thus, so-called parametric *null imposed* critical values in Andrews and Cheng (2012a) for t-, Quasi-Likelihood Ratio and Wald statistics do not play a role here.

E.1.2 Identification Category Selection Type 1

The LF critical value does not exploit data related information that may point toward a particular identification case. The ICS procedure uses the sample to choose between $\sqrt{n}\beta_n \to b$ when $||b|| < \infty$ (weak and non-identification) and $||b|| = \infty$ (semi-strong and strong identification).

Recall the statistic $\mathcal{A}_n$ in (A.6). Now let $\{\kappa_n\}$ be a sequence of positive constants, with $\kappa_n \to \infty$ and $\kappa_n = o(n^{1/2})$. The case $||b|| < \infty$ is selected when $\mathcal{A}_n \leq \kappa_n$, else $||b|| = \infty$ is selected. Now define the type 1 ICS [ICS-1] critical value: $c_{1-\alpha,n}^{(ICS-1)}(\lambda) = c_{1-\alpha}^{(LF)}(\lambda)$ if $\mathcal{A}_n \leq \kappa_n$, else $c_{1-\alpha,n}^{(ICS-1)}(\lambda) = \chi^2_{1-\alpha}$ if $\mathcal{A}_n > \kappa_n$.

$$c_{1-\alpha,n}^{(ICS-1)}(\lambda) = \begin{cases} c_{1-\alpha}^{(LF)}(\lambda) & \text{if } \mathcal{A}_n \leq \kappa_n \\ \chi^2_{1-\alpha} & \text{if } \mathcal{A}_n > \kappa_n \end{cases}.$$

See the remark following Theorem 6.1, and Andrews and Cheng (2012a, p. 2191), for intuition on $c_{1-\alpha,n}^{(ICS-1)}(\lambda)$. Briefly: only when $\sqrt{n}||\beta_n|| \to \infty$ faster than $\kappa_n \to \infty$ will the chi-squared based critical value be chosen asymptotically with probability approaching one since then $\mathcal{A}_n/\kappa_n \xrightarrow{p} \infty$. Thus, a high bar must be passed in order for the strong identification case to be selected. In every other case the LF value is chosen, which is always asymptotically correct.

E.1.3 Identification Category Selection Type 2

Let $s : [0, \infty) \to [0, 1]$ be a continuous function, $s(x)$ is non-increasing in x, $s(0) = 1$, and $s(x) \to 0$ as $x \to \infty$. An example is $s(x) = \exp\{-cx\}$ for some $c > 0$. Let $(\Delta_1, \Delta_2) \geq 0$ and $\kappa > 0$ be user selected numbers. Define

$$\begin{aligned} c_1(\lambda) &= c_{1-\alpha}^{(LF)}(\lambda) + \Delta_1 \\ c_2(\lambda) &= \chi^2_{1-\alpha} + \Delta_2 + (c_{1-\alpha}^{(LF)}(\lambda) - \chi^2_{1-\alpha} + \Delta_1 - \Delta_2)s(\mathcal{A}_n - \kappa). \end{aligned}$$

The type 2 ICS [ICS-2] critical value is

$$c_{1-\alpha,n}^{(ICS-2)}(\lambda) = \begin{cases} c_1(\lambda) & \text{if } \mathcal{A}_n \leq \kappa \\ c_2(\lambda) & \text{if } \mathcal{A}_n > \kappa \end{cases}.$$

The construction allows for a smooth transition between identification cases, and allows for a non-diverging threshold. The latter necessitates the tuning parameters (Δ_1, Δ_2) which promote a correct asymptotic size. See also Andrews and Barwick (2012) for a related method.

See Andrews and Cheng (2012a, p. 2193) for details on determining appropriate choices for $(\Delta_1, \Delta_2, \kappa)$, and see Appendix D above.

E.2 Asymptotics for Robust Critical Values

Let $c^{(\cdot)}_{1-\alpha,n}(\lambda)$ denote the LF, ICS-1 or ICS-2 plug-in robust critical value. Conditions leading to critical value asymptotics follow, and are presented in Andrews and Cheng (2012a, Section 5) and Andrews and Cheng (2013a, Section 5.5).

Assumption 7 (critical value)**.** *If $c^{(\cdot)}_{1-\alpha,n}(\lambda)$ is (i) LF, (ii) ICS-1, or (iii) ICS-2, then assume respectively that Andrews and Cheng's (2012a) Assumption (i) LF, (ii) K and V3, or (iiii) Rob2 holds.*

Let F_γ be the distribution function of W_t under some $\gamma \in \Gamma^*$, where Γ^* is the true parameter space in (E.32). Let P_γ denote probability under F_γ. For any critical value $c^{(\cdot)}_{1-\alpha,n}(\lambda)$ and each λ the asymptotic size of the test is the maximum rejection probability over γ such that the null is true:

$$AsySz(\lambda) = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma \left(\mathcal{T}_n(\lambda) > c^{(\cdot)}_{1-\alpha,n}(\lambda) | H_0 \right).$$

Proofs are presented in Appendix E.4.

Theorem E.1. *Under Assumptions 1-2, 4 and 7 and H_0, the LF, ICS-1 and ICS-2 $c^{(\cdot)}_{1-\alpha,n}(\lambda)$ satisfy $AsySz(\lambda) = \alpha$.*

E.3 Computation of $c^{(\cdot)}_{1-\alpha,n}(\lambda)$

Steps 1-4 of the wild bootstrap procedure outlined in Section 6.2 of the main paper carries over verbatim.

Step 5 is as follows. Repeat Steps 1-4 $\mathcal{M}$ times resulting in a sequence of independent draws $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda, h)\}_{j=1}^{\mathcal{M}}$. Define order statistics $\hat{\mathcal{T}}^*_{\psi,n,[1]}(\lambda, h) \leq \hat{\mathcal{T}}^*_{\psi,n,[2]}(\lambda, h) \leq \cdots$. The critical value approximation is $\hat{c}^*_{1-\alpha,n,\mathcal{M}}(\lambda, h) \equiv \hat{\mathcal{T}}^*_{\psi,n,[(1-\alpha)\mathcal{M}]}(\lambda, h)$, which is consistent for the asymptotic critical value $c_{1-\alpha}(\lambda, h)$.

Theorem E.2. *Let the true value $\sigma^2 \equiv E[\epsilon_t^2] \in \mathfrak{S}^*$, where the true parameter space $\mathfrak{S}^*$ is a compact subset of $(0, \infty)$. Let $\mathcal{M} = \mathcal{M}_n \to \infty$ as $n \to \infty$. Under Assumptions 1-2, 4 and 7, $\hat{c}^*_{1-\alpha,n,\mathcal{M}_n}(\lambda, h) \xrightarrow{p} c_{1-\alpha}(\lambda, h)$ for each $h \in \mathfrak{H}$ and $\lambda \in \Lambda$.*

E.4 Proofs

Proof of Theorem E.1.

Step 1 (LF). The proof for the LF critical value $c_{1-\alpha}^{(LF)} = \max\{\sup_{h\in\mathfrak{H}}\{c_{1-\alpha}(\lambda,h)\},\chi^2_{1-\alpha}\}$ is identical to arguments in Andrews and Cheng (2012b, Appendix B: proof of Theorem 5.1). We verify the conditions of Lemma 2.1 in Andrews and Cheng (2012a) below. An application of their Lemma 2.1 to the asymptotic size for $\mathcal{T}_\psi(\lambda,h)$, and Theorem 4.2.a, yields

$$AsySz(\lambda) = \max\left\{\sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}^{(LF)}\right), P\left(\mathcal{T}(\lambda) > c_{1-\alpha}^{(LF)}\right)\right\}.$$

If $c_{1-\alpha}^{(LF)} = \chi^2_{1-\alpha}$ then by the definition of $c_{1-\alpha}(\lambda,h)$:

$$\sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}^{(LF)}\right) \leq \sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}(\lambda,h)\right) = \alpha,$$

hence $AsySz(\lambda)$ is:

$$\max\left\{\sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > \chi^2_{1-\alpha}\right), P\left(\mathcal{T}(\lambda) > \chi^2_{1-\alpha}\right)\right\} = \max\left\{\sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > \chi^2_{1-\alpha}\right), \alpha\right\} = \alpha.$$

Conversely, if $c_{1-\alpha}^{(LF)} = \sup_{h\in\mathfrak{H}}\{c_{1-\alpha}(\lambda,h)\}$ then

$$\sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}^{(LF)}\right) = \sup_{h\in\mathfrak{H}} P\left(\mathcal{T}_\psi(\lambda,h) > \sup_{h\in\mathfrak{H}}\{c_{1-\alpha}(\lambda,h)\}\right) = \alpha,$$

and $P(\mathcal{T}(\lambda) > c_{1-\alpha}^{(LF)}) \leq \alpha$ hence again $AsySz(\lambda) = \alpha$.

It remains to verify the conditions of Lemma 2.1 in Andrews and Cheng (2012a). We must show their Assumption ACP holds, parts (i)-(iv). Recall $\{\gamma_n\}$ is a sequence of true parameters $\gamma_n \equiv (\theta_n,\phi_0)$ under local drift which fully determine the joint distribution of the data $[y_t, y_{t-1}, ..., y_{t-p}]'$. The limiting true value is $\gamma_0 \equiv (\theta_0,\phi_0)$. By Theorem 4.2, $P_{\gamma_n}(\mathcal{T}_n(\lambda) > c_{1-\alpha}^{(LF)})$ $\rightarrow P(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}^{(LF)})$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$, and $P_{\gamma_n}(\mathcal{T}_n(\lambda) > c_{1-\alpha}^{(LF)}) \rightarrow P(\mathcal{T}(\lambda) > c_{1-\alpha}^{(LF)})$ under $\mathcal{C}(ii,\omega_0)$. Hence Assumption ACP.i,ii,iii hold. Assumption ACP.iv holds under true parameter space Assumption 1.e, because the latter is identically Assumption STAR4 in Andrews and Cheng (2013a), cf. Andrews and Cheng (2013b, Section 15.7).

Step 2 (ICS-1, ICS-2). Theorem 5.1 implies the ICS statistic satisfies $\mathcal{A}_n = O_p(1)$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$. Under $\mathcal{C}(ii,\omega_0)$ we have $\mathcal{A}_n \xrightarrow{p} \infty$, and if $\beta_0 \neq 0$ then $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} \infty$ where by supposition $\kappa_n \rightarrow \infty$ and $\kappa_n = o(\sqrt{n})$. Now invoke Theorem 4.2 to deduce $P_{\gamma_n}(\mathcal{T}_n(\lambda) > c_{1-\alpha,n}^{(ICS-1)}(\lambda)) \rightarrow P(\mathcal{T}_\psi(\lambda,h) > c_{1-\alpha}^{(LF)})$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$, and $P_{\gamma_n}(\mathcal{T}_n(\lambda) > c_{1-\alpha,n}^{(ICS-1)}) \rightarrow$

$P(\mathcal{T}(\lambda) > \chi^2_{1-\alpha})$ under $\mathcal{C}(ii, \omega_0)$ if $\beta_0 \neq 0$. Hence Assumption ACP.i,ii,iii in Andrews and Cheng (2012a) hold. Their Assumption ACP.iv holds by Step 1. Arguments in Andrews and Cheng (2012b, p. 56-58) now carry over to prove the ICS-1 and ICS-2 claims. $\mathcal{QED}$.

Proof of Theorem E.2. By Step 1 in the proof of Theorem 6.2:

$$\left\{\hat{\mathcal{T}}^*_{\psi,n}(\lambda,h) : \lambda \in \Lambda\right\} \Rightarrow^p \left\{\left(\frac{\mathfrak{T}_\psi(\pi^*(b),\lambda,b)}{\bar{v}(\pi^*(b),\lambda,b)}\right)^2 : \lambda \in \Lambda\right\} = \left\{\mathcal{T}_\psi(\lambda,h) : \lambda \in \Lambda\right\}, \tag{E.34}$$

the Theorem 5.2 null limit process under weak identification.

Define quantile functions

$$\begin{aligned}
\hat{F}^{-1}_{n,\lambda}(u|\cdot) &\equiv \inf\left\{c \geq 0 : P(\hat{\mathcal{T}}^*_{\psi,n,1} \leq c) \geq u|\cdot\right\} \\
F^{-1}_{n,\lambda}(u) &\equiv \inf\left\{c \geq 0 : P(\mathcal{T}_n(\lambda) \leq c) \geq u\right\} \\
F^{-1}_{\lambda,h}(u) &\equiv \inf\left\{c \geq 0 : P(\mathcal{T}_\psi(\lambda,h) \leq c) \geq u\right\}
\end{aligned}$$

By Theorem 5.2.a, $\{\mathcal{T}_n(\lambda) : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda,h) : \Lambda\}$ under H_0 and $\mathcal{C}(i,b)$ with $||b|| < \infty$. Weak convergence implies convergence in finite dimensional distribution. By the construction of distribution convergence it therefore follows that $F^{-1}_{n,\lambda}(u) \rightarrow F^{-1}_{\lambda,h}(u)$.

Now operate conditionally on the sample $\mathfrak{W}_n$. By weak convergence in probability (E.34), $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)\}_{j=1}^{\mathcal{M}}$ is a sequence of iid draws from $\{\mathcal{T}_\psi(\lambda,h) : \Lambda\}$, asymptotically with probability approaching one with respect to the draw $\mathfrak{W}_n \equiv \{(y_t,x_t)\}_{t=1}^n$. Therefore $\mathcal{T}_n(\lambda)$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$, and $\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h)$ have the same weak limits in probability under H_0. Since $\mathcal{T}_n(\lambda)$, and $\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)$ conditionally on $\mathfrak{W}_n$ have the same weak limits in probability under H_0, it follows that (see Gine and Zinn, 1990, Section 3, eq's (3.4) and (3.5))

$$\sup_{c \geq 0}\left|P(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) \leq c|\mathfrak{W}_n) - F_{n,\lambda}(c)\right| \xrightarrow{p} 0 \ \forall \lambda \in \Lambda.$$

Therefore, by construction of convergence of probability measures (see, e.g., Chapt. 21 in van der Vaart, 1998):

$$\sup_{u \in [0,1]}\left|\hat{F}^{-1}_{n,\lambda}(u|\mathfrak{W}_n) - F^{-1}_{n,\lambda}(u)\right| \xrightarrow{p} 0 \ \forall \lambda \in \Lambda.$$

Moreover, by independence and $\mathcal{M}_n \rightarrow \infty$, the bootstrapped critical value $\hat{c}^*_{1-\alpha,n,\mathcal{M}_n}(\lambda,h) \equiv \hat{\mathcal{T}}^*_{\psi,n,[(1-\alpha)\mathcal{M}_n]}(\lambda,h)$ is a central order statistic of a (conditionally) iid random variable, hence pointwise on Λ:

$$\left|\hat{c}^*_{1-\alpha,n,\mathcal{M}_n}(\lambda,h) - \hat{F}^{-1}_{n,\lambda}(1-\alpha|\mathfrak{W}_n)\right| \xrightarrow{p} 0.$$

See, e.g., Galambos (1987), for a classic treatment of order statistics. Now combine

$$\begin{aligned}
&\left|\hat{c}^*_{1-\alpha,n,\mathcal{M}_n}(\lambda,h) - \hat{F}^{-1}_{n,\lambda}(1-\alpha|\mathfrak{W}_n)\right| \overset{p}{\to} 0 \\
&\left|\hat{F}^{-1}_{n,\lambda}(1-\alpha|\mathfrak{W}_n) - F^{-1}_{n,\lambda}(1-\alpha)\right| \overset{p}{\to} 0 \\
&F^{-1}_{n,\lambda}(1-\alpha) \to F^{-1}_{\lambda,h}(1-\alpha)
\end{aligned}$$

to yield

$$|\hat{c}^*_{1-\alpha,n,\mathcal{M}_n}(\lambda,h) - F^{-1}_{\lambda,h}(1-\alpha) \overset{p}{\to} 0.$$

By definition $c_{1-\alpha}(\lambda,h) = F^{-1}_{\lambda,h}(1-\alpha)$ hence the proof is complete. $\mathcal{QED}$.

F Example: STAR Model (Assumptions 3, 4, 5

We discuss Assumptions 3, 4 and 5 for a simple STAR model. The data generating properties in Assumption 1 along with the minimization conditions for the process $\{\xi_\psi(\pi,b) : \pi \in \Pi\}$ under Assumption 2 are treated at length in Andrews and Cheng (2013b, Section 7) and Andrews and Cheng (2013b, Appendix E).

The model is a simplified Exponential STAR(1) for ease of exposition (cf. Terasvirta, 1994):

$$y_t = \beta_0 y_{t-1}\exp\left\{-\pi_0 y^2_{t-1}\right\} + \epsilon_t \text{ where } \pi_0 > 0, \text{ hence } g(y_{t-1},\pi_0) = y_{t-1}\exp\left\{-\pi_0 y^2_{t-1}\right\}.$$

Assume y_t is strictly stationary, $E|y_t|^r < \infty$ for some $r > 6$, and $\mathcal{F}_t \equiv \sigma(y_\tau : \tau \leq t)$ is strictly increasing $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ $\forall t$. ϵ_t has a (non-degenerate) continuous distribution on $\mathbb{R}$ $\forall t$, $E[\epsilon_t] = 0$ and $\pi_0 \in \Pi \subset (0,\infty)$. Hence y_t has a (non-degenerate) continuous distribution. Assume $E[\epsilon^2_t|y_{t-1}]$ $= \sigma^2_0$ $a.s.$ for some finite $\sigma^2_0 > 0$.

Let the compact nuisance parameter space be $\Lambda \subset \mathbb{R}/0$. We omit $\lambda = 0$ because $F(0 \times y_{t-1})$ $= F(0)$ is a constant and cannot therefore reveal model misspecification (cf. Bierens, 1990; Stinchcombe and White, 1998).

We first define some useful components:

$$d_{\psi,t}(\pi) \equiv g(y_{t-1},\pi_0) = y_{t-1}\exp\left\{-\pi y^2_{t-1}\right\}$$

$$d_{\theta,t}(\omega,\pi) \equiv \left[y_{t-1}\exp\left\{-\pi y^2_{t-1}\right\}, -\omega y^3_{t-1}\exp\left\{-\pi y^2_{t-1}\right\}\right]'$$

$$\mathcal{D}_\psi(\pi) \equiv -E\left[y^2_{t-1}\exp\left\{-2\pi y^2_{t-1}\right\}\right] = -\mathcal{H}_\psi(\pi)$$

$$\mathcal{H}_\psi(\pi) \equiv E\left[y_{t-1}^2 \exp\left\{-2\pi y_{t-1}^2\right\}\right] > 0 \ \forall \pi \in \Pi$$

$$\mathcal{H}_\theta(\omega,\pi) \equiv E\left[d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)'\right] = \begin{bmatrix} E\left[y_{t-1}^2 \exp\left\{-2\pi y_{t-1}^2\right\}\right] & -\omega E\left[y_{t-1}^4 \exp\left\{-2\pi y_{t-1}^2\right\}\right] \\ -\omega E\left[y_{t-1}^4 \exp\left\{-2\pi y_{t-1}^2\right\}\right] & \omega^2 E\left[y_{t-1}^6 \exp\left\{-2\pi y_{t-1}^2\right\}\right] \end{bmatrix}$$

$$\mathfrak{b}_\psi(\pi,\lambda) \equiv E\left[F\left(\lambda \mathcal{W}(y_{t-1})\right) y_{t-1} \exp\left\{-\pi y_{t-1}^2\right\}\right]$$

$$\mathfrak{b}_\theta(\omega,\pi,\lambda) \equiv E\left[F\left(\lambda \mathcal{W}(y_{t-1})\right)\left[y_{t-1} \exp\left\{-\pi y_{t-1}^2\right\}, -\omega y_{t-1}^3 \exp\left\{-\pi y_{t-1}^2\right\}\right]'\right]$$

$$\mathcal{K}_{\psi,t}(\pi,\lambda) \equiv F\left(\lambda' W(y_{t-1})\right) - \mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)d_{\psi,t}(\pi)$$

$$\mathcal{K}_{\theta,t}(\lambda) \equiv F\left(\lambda' \mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}(\beta_n/\left\|\beta_n\right\|,\pi_0)$$

Under the stated conditions:

$$\inf_{\pi\in\Pi} \mathcal{H}_\psi(\pi) > 0.$$

Now write $\Pi = [\pi_L, \pi_H]$ for some $0 < \pi_L < \pi_H < \infty$. Similarly, for $r = [r_1, r_2]'$,

$$\inf_{r'r=1} \inf_{\pi\in\Pi} r'\mathcal{H}_\theta(\omega,\pi)r > 0,$$

because under the stated conditions:

$$\begin{aligned} \inf_{r'r=1} \inf_{\pi\in\Pi} r'\mathcal{H}_\theta(\omega,\pi)r &= \inf_{r'r=1} \inf_{\pi\in\Pi} E\left[\left(r_1 y_{t-1} + r_2 y_{t-1}^3\right)^2 \exp\left\{-2\pi y_{t-1}^2\right\}\right] \\ &= \inf_{r'r=1} E\left[y_{t-1}^2\left(r_1 + r_2 y_{t-1}^2\right)^2 \exp\left\{-2\pi_H y_{t-1}^2\right\}\right] \\ &= 0 \end{aligned}$$

if and only if $r_1 + r_2 y_{t-1}^2 = 0$ *a.s.* for some $r'r = 1$. The condition $r_1 + r_2 y_{t-1}^2 = 0$ *a.s.* is ruled out due to $r'r = 1$ and y_{t-1} having a non-degenerate continuous distribution on $\mathbb{R}$.

F.1 Assumption 3

We tackle part (a); part (b) is similar. First, we have:

$$\kappa_t(\omega,\pi) \equiv [\mu(y_{t-1}), y_{t-1} \exp\left\{-\pi y_{t-1}^2\right\}, -\omega y_{t-1}^3 \exp\left\{-\pi y_{t-1}^2\right\}]' \in \mathbb{R}^5.$$

Note that ω is a scalar because β is, hence $\omega^2 = 1$ implies $\omega \in [-1, 1]$. It therefore suffices to show that there exists a Borel measurable function $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$ (recall $k_x = 1$) such that:

$$\inf_{\omega \in [-1,1], \pi \in \Pi} \left\{ \inf_{r \in \mathbb{R}^5 : r'r=1} E\left[(r'\kappa_t(\omega, \pi))^2 \right] \right\} > 0.$$

Suppose the contrary holds. Then for every Borel measurable μ, there exist $\alpha \in \mathbb{R}^3$, $\alpha'\alpha = 1$, and some $\pi \in \Pi$ such that:

$$\alpha_1 \mu(y_t) + \alpha_2 y_t \exp\left\{-\pi y_t^2\right\} + \alpha_3 y_t^3 \exp\left\{-\pi y_t^2\right\} = 0 \ a.s. \ \forall t.$$

The key idea is to find a μ that leads to a contradiction of the primitive assumptions. Such μ are easily found: consider $\mu(y_t) = y_t$. Then

$$\alpha_1 y_t + \alpha_2 y_t \exp\left\{-\pi y_t^2\right\} + \alpha_3 y_t^3 \exp\left\{-\pi y_t^2\right\} = 0 \ a.s. \ \forall t. \tag{F.35}$$

For any fixed $\alpha \in \mathbb{R}^3$, $\alpha'\alpha = 1$, and $0 < \pi < \infty$, (F.35) can only hold if y_t has a degenerate distribution and $\mathcal{F}_t = \mathcal{F}_{t+1}$, which contradicts distribution nondegeneracy and $\mathcal{F}_t \subset \mathcal{F}_{t+1}$.

F.2 Assumption 4

We now discuss Assumption 4. The assumption cannot generally be verified, which is precisely why is must be assumed (cf. Bierens, 1990, p. 1449). We do, however, present some refinements revealing greater details behind test statistic variance degeneracy.

F.2.1 General Test Weight

We only discuss the simplest case: case (a) under strong identification $\mathcal{C}(ii, \omega_0)$. This gives the basic intuition behind the requirement of the assumption. Write $\epsilon_t(\theta) \equiv y_t - \beta y_{t-1} \exp\left\{-\pi y_{t-1}^2\right\}$. The assumption requires $v^2(\theta_0, \lambda) > 0 \ \forall \lambda \in \Lambda$ where

$$v^2(\theta, \lambda) = E\left[\epsilon_t^2(\theta) \left\{ F\left(\lambda \mathcal{W}(y_{t-1})\right) - \mathfrak{b}_\theta(\omega(\beta), \pi, \lambda)' \mathcal{H}_\theta^{-1}(\omega(\beta), \pi) d_{\theta,t}(\omega(\beta), \pi) \right\}^2 \right].$$

Define

$$v^2(\omega, \pi, \lambda) \equiv E\left[\epsilon_t^2(\psi_0, \pi) \left\{ F\left(\lambda \mathcal{W}(y_{t-1})\right) - \mathfrak{b}_\theta(\omega, \pi, \lambda)' \mathcal{H}_\theta^{-1}(\omega, \pi) d_{\theta,t}(\omega, \pi) \right\}^2 \right].$$

By Lemma B.12, under Assumptions 1.a(i) and 3 we know $\inf_{\omega'\omega=1, \pi \in \Pi} v^2(\omega, \pi, \lambda) = 0$ only on a subset $S^* \subset \Lambda$ with measure zero. See Bierens (1990, Lemma 2) for an original treatment of this property. Hence $v^2(\theta_0, \lambda) > 0 \; \forall \lambda \in \Lambda / S^*$ where S^* is countable. Further, Theorem 4 in Hill (2008) extends to any valid $F(\cdot)$ considered here. Hence $S^* \subseteq S$ where S is the countable set on which $E[\epsilon_t F(\lambda \mathcal{W}(y_{t-1}))] = 0$ under H_1. That is, any λ such that $v^2(\theta_0, \lambda) = 0$ actually has a two-fold failure since also $E[\epsilon_t F(\lambda \mathcal{W}(y_{t-1}))]$ fails to detect misspecification. Although we know $S^* \subseteq S$, this does not provide a context in which we can deduce $S^* = \varnothing$ such that Assumption 4 holds. Generally the sets S^* and S depend on the underlying joint distribution, but deriving the exact contents of either set, let alone proving $S^* = \varnothing$, is evidently not feasible. The only way either set can be viewed is by simulation study (see, e.g., Bierens, 1990; Hill, 2013).

F.2.2 Vector Test Weight

We can go somewhat further by studying a specific class of *vector* test weights that never fail to reveal model misspecification. Unfortunately, even here we cannot prove the appropriate asymptotic variance is positive definite for all nuisance parameters $\lambda \in \Lambda$ due to the vector nature of the moment condition.

We first derive the vector test weight, and the appropriate asymptotic variance matrix for the implied vector sample moment condition. We then show that although the vector test weight reveals model misspecification for all $\lambda \in \Lambda$, the asymptotic variance need not be positive definite for all $\lambda \in \Lambda$.

Moment Condition Define

$$\xi^{(+)} \equiv \arg\sup_{\lambda \in \Lambda} \frac{\partial}{\partial \lambda} E[\epsilon_t F(\lambda \mathcal{W}(y_{t-1}))] \text{ and } F'(u) \equiv \frac{\partial}{\partial u} F(u).$$

Hill (2013) shows that by stacking the test weights

$$w_t(\lambda) \equiv \left[F(\lambda \mathcal{W}(y_{t-1})), y_{t-1} F'\left(\xi^{(+)} \mathcal{W}(y_{t-1})\right) \right]',$$

a perfectly revealing test weight is achieved in the sense that:

$$\text{under } H_1 : E\left[\epsilon_t w_t(\lambda)\right] \neq 0 \; a.s. \; \forall \lambda \in \Lambda / S \text{ where } S = \{0\} \text{ or } \varnothing.$$

We assume $0 \notin \Lambda$ hence S is empty. A similar result applies if we use $\xi^{(-)} \equiv \arg\inf_{\lambda \in \Lambda} (\partial / \partial \lambda) E[\epsilon_t F(\lambda \mathcal{W}(y_{t-1}))]$ or use both $y_{t-1} F'\left(\xi^{(+)} \mathcal{W}(y_{t-1})\right)]$ and $y_{t-1} F'\left(\xi^{(-)} \mathcal{W}(y_{t-1})\right)]$ in $w_t(\lambda)$. See Hill (2013, Section 2.2, Theorem A.1).

Asymptotic Variance Matrix Using ideas in the main paper, it is straightforward to show that the appropriate scale for the standardized sample vector moment condition

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda) \in \mathbb{R}^2$$

is the matrix

$$\begin{aligned}\hat{\mathcal{V}}_n(\hat{\theta}_n,\lambda) \;\; &\equiv \;\; \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\hat{\theta}_n)\left\{w_t(\lambda) - \hat{\mathfrak{b}}_{\theta,n}(\omega(\hat{\beta}_n),\hat{\pi}_n,\lambda)'\widehat{\mathcal{H}}_n^{-1}d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)\right\} \\ &\qquad\qquad \times\left\{w_t(\lambda) - \hat{\mathfrak{b}}_{\theta,n}(\omega(\hat{\beta}_n),\hat{\pi}_n,\lambda)'\widehat{\mathcal{H}}_n^{-1}d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)\right\}',\end{aligned}$$

where

$$\hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n}w_t(\lambda)d_{\theta,t}(\omega,\pi).$$

Notice the only differences with $\hat{\mathcal{V}}_n(\hat{\theta}_n,\lambda)$ here and $\hat{v}_n^2(\hat{\theta}_n,\lambda)$ in the main paper are (i) $\hat{\mathcal{V}}_n(\hat{\theta}_n,\lambda)$ is a matrix; and (ii) $\hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda)$ is defined using $w_t(\lambda)$ instead of just $F(\lambda\mathcal{W}(y_{t-1}))$.

Write compactly:

$$\mathfrak{b}_\theta(\lambda) = \mathfrak{b}_\theta(\omega(\beta_0),\pi_0,\lambda),\;\; \mathcal{H}_\theta = \mathcal{H}_\theta(\omega(\beta_0),\pi_0),\;\; d_{\theta,t} = d_{\theta,t}(\omega(\beta_0),\pi_0).$$

The probability limit of $\hat{\mathcal{V}}_n(\hat{\theta}_n,\lambda)$ is

$$\mathcal{V}(\theta_0,\lambda) = E\;\left[\epsilon_t^2\left\{w_t(\lambda) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\right\}\left\{w_t(\lambda) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\right\}'\right].$$

Non-Positive Definiteness For fixed λ if $r_\lambda'\mathcal{V}(\theta_0,\lambda)r_\lambda = 0$ for some $r_\lambda'r_\lambda = 1$, then:

$$r_\lambda'w_t(\lambda) = r_\lambda'\mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\;a.s.$$

Now use $E[\epsilon_t d_{\theta,t}] = 0$ under Assumption 1.a(ii) to yield:

$$E\left[\epsilon_t r_\lambda'w_t(\lambda)\right] = r_\lambda'\mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}E\left[\epsilon_t d_{\theta,t}\right] = 0.$$

Therefore, for λ such that $\mathcal{V}(\theta_0,\lambda)$ is non-positive definite, a failed moment condition $E[\epsilon_t r_\lambda'w_t(\lambda)] = 0$ occurs under H_1 for some r_λ despite $E[\epsilon_t w_t(\lambda)] \neq 0\ \forall\lambda$. Unfortunately there is nothing that precludes $r_\lambda'\mathcal{V}(\theta_0,\lambda)r_\lambda = 0$ for some λ and $r_\lambda'r_\lambda = 1$: we cannot prove $\inf_{r'r=1} r'\mathcal{V}(\theta_0,\lambda)r > 0\ \forall\lambda \in \Lambda$. Thus, since it is easily shown that $r'w_t(\lambda)$ for any $r'r = 1$ satisfies the required test weight properties, we can only say $E[\epsilon_t r'w_t(\lambda)] \neq 0$ under H_1 $\forall\lambda \in \Lambda/S_r$ where S_r has measure zero.

This is a key shortcoming because a quadratic-type test statistic

$$\mathcal{T}_n(\lambda) \equiv \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda)\right)^{\prime}\hat{\mathcal{V}}_n^{-1}(\hat{\theta}_n,\lambda)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda)\right)$$

is just the inner product of linearly combined sample moments:

$$\mathcal{T}_n(\lambda) = n\left(\widehat{\mathcal{A}}_n(\hat{\theta}_n,\lambda)^{\prime}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda)\right)^{\prime}\left(\widehat{\mathcal{A}}_n(\hat{\theta}_n,\lambda)^{\prime}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda)\right)$$

where $\widehat{\mathcal{A}}_n(\hat{\theta}_n,\lambda)\widehat{\mathcal{A}}_n(\hat{\theta}_n,\lambda)^{\prime} = \hat{\mathcal{V}}_n^{-1}(\hat{\theta}_n,\lambda)$ is assumed to exist $a.s.$ for each n. If $\mathcal{V}(\theta_0,\lambda)$ is non-positive definite at λ then $E[\epsilon_t r_\lambda^{\prime} w_t(\lambda)] = 0$ for some $\lambda \in \Lambda$ and $r_\lambda \neq 0$, hence $\widehat{\mathcal{A}}_n(\hat{\theta}_n,\lambda)^{\prime}1/n\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda)$ $\xrightarrow{p} 0$ under H_1 is possible even though $1/n\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_t(\lambda) \overset{p}{\nrightarrow} 0$ $\forall\lambda \in \Lambda$ under H_1. Of course, by non-positive definiteness, $\hat{\mathcal{V}}_n^{-1}(\hat{\theta}_n,\lambda)$ does not have a probability limit and therefore $\mathcal{T}_n(\lambda)$ does not have a non-degenerate limit distribution under H_0.

Alternative Approach A better approach is therefore to by-pass standardization (and therefore standard asymptotics) altogether. One path is to use the test statistic

$$\max_{i=1,2}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_{i,t}(\lambda)\right| \text{ where } w_t(\lambda) = [w_{1,t}(\lambda), w_{2,t}(\lambda)]^{\prime},$$

or a standardize version of it. Under the null:

$$\left\{\max_{i=1,2}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)w_{i,t}(\lambda)\right| : \lambda \in \Lambda\right\} \Rightarrow^{*} \left\{\max_{i=1,2}|\mathcal{Z}_i(\lambda)| : \lambda \in \Lambda\right\}$$

where $\{[\mathcal{Z}_1(\lambda), \mathcal{Z}_2(\lambda)] : \lambda \in \Lambda\}$ is zero mean Gaussian process with *almost surely* bounded and uniformly continuous sample paths. This limit process can be easily bootstrapped by multiplier (wild) bootstrap. We leave this idea for future consideration.

F.3 Assumption 5

Only (a) needs discussion since under (b) the analyst sets the ICS-1 threshold sequence $\{\kappa_n\}$ to satisfy $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$.

Recall $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda,h) \leq c)$ where $\{\mathcal{T}_\psi(\lambda,h) : \lambda \in \Lambda\}$ is the asymptotic null process under weak identification. Under (a) we need $\mathcal{F}_{\lambda,h}(\cdot)$ to be continuous $a.e.$ on $[0,\infty)$, $\forall h \in \mathfrak{H}$.

Using the notation of Section 4 in the main paper, recall

$$\tau_\beta(\pi, b) \equiv -\mathcal{S}_\beta \mathcal{H}_\psi^{-1}(\pi) \{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\} \text{ where } \mathcal{S}_\beta \equiv [1, 0],$$

and

$$\begin{aligned} \mathfrak{T}_\psi(\pi, \lambda, b) &\equiv \mathfrak{Z}_\psi(\pi, \lambda) + \mathfrak{b}_\psi(\pi, \lambda)' \left\{\mathcal{H}_\psi^{-1}(\pi)\mathcal{D}_\psi(\pi) b + \left[b, 0'_{k_\beta}\right]'\right\} \\ &\quad + \mathfrak{b}_\psi(\pi, \lambda)' \mathcal{H}_\psi^{-1}(\pi) E\left[d_{\psi,t}(\pi)\{g(y_{t-1}, \pi_0) - g(y_{t-1}, \pi)\}'\right] b \\ &\quad + E\left[\mathcal{K}_{\psi,t}(\pi, \lambda)\{g(y_{t-1}, \pi_0) - g(y_{t-1}, \pi)\}'\right] b \\ &\equiv \mathfrak{Z}_\psi(\pi, \lambda) + \mathcal{W}_\psi(\pi, \lambda), \end{aligned}$$

say, and

$$v^2(\omega, \pi, \lambda) \equiv E\left[\epsilon_t^2(\psi_0, \pi)\left\{F\left(\lambda \mathcal{W}(y_{t-1})\right) - \mathfrak{b}_\theta(\omega, \pi, \lambda)' \mathcal{H}_\theta^{-1}(\omega, \pi) d_{\theta,t}(\omega, \pi)\right\}^2\right]$$

$$\bar{v}^2(\pi, \lambda, b) \equiv v^2(\omega^*(\pi, b), \pi, \lambda) \text{ where } \omega^*(\pi, b) \equiv \tau_\beta(\pi, b) / \|\tau_\beta(\pi, b)\|.$$

Then

$$\mathcal{T}_\psi(\pi, \lambda, b) \equiv \frac{\mathfrak{T}_\psi^2(\pi, \lambda, b)}{\bar{v}^2(\pi, \lambda, b)} \text{ and } \mathcal{T}_\psi(\lambda, b) \equiv \mathcal{T}_\psi(\pi^*(b), \lambda, b)$$

where

$$\begin{aligned} \pi^*(b) &= \arg\inf_{\pi \in \Pi} \xi_\psi(\pi, b) \\ &\equiv \arg\inf_{\pi \in \Pi} \left\{-\frac{1}{2}\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\}' \mathcal{H}_\psi^{-1}(\pi)\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\}\right\}. \end{aligned}$$

F.3.1 Numerator $\mathfrak{T}_\psi^2(\pi, \lambda, b)$

The only stochastic component of $\mathfrak{T}_\psi(\pi, \lambda, b) = \mathfrak{Z}_\psi(\pi, \lambda) + \mathcal{W}_\psi(\pi, \lambda)$ is $\mathfrak{Z}_\psi(\pi, \lambda)$. Recall by Lemma B.9 that $\mathfrak{Z}_\psi(\pi, \lambda)$ is a limit process under H_0

$$\left\{\frac{1}{\sqrt{n}} \sum_{t=1}^n \epsilon_t \mathcal{K}_{\psi,t}(\pi, \lambda) : \Pi, \Lambda\right\} \Rightarrow^* \{\mathfrak{Z}_\psi(\pi, \lambda) : \Pi, \Lambda\}$$

where $\{\mathfrak{Z}_\psi(\pi, \lambda) : \Pi, \Lambda\}$ is a zero mean Gaussian process with *almost surely* uniformly continuous, and bounded, sample paths, and covariance kernel $\sigma_0^2 E[\mathcal{K}_{\psi,t}(\pi, \lambda)\mathcal{K}_{\psi,t}(\tilde{\pi}, \tilde{\lambda})]$. In view of the remaining components in $\mathfrak{T}_\psi(\pi,\lambda,b)$, it follows easily that $\{\mathfrak{T}_\psi(\pi,\lambda,b) : \Pi,\Lambda\}$ is a Gaussian process with continuous and bounded sample paths.

Next, stochastic $\pi^*(b) = \arg\inf_{\pi\in\Pi} \xi_\psi(\pi,b)$ minimizes

$$\xi_\psi(\pi,b) \equiv -\frac{1}{2}\left\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi)b\right\}' \mathcal{H}_\psi^{-1}(\pi)\left\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi)b\right\}.$$

The only stochastic component here is $\mathcal{G}_\psi(\pi)$. By Lemma B.1 and continuity,

$$\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) \equiv -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_{0,n})d_{\psi,t}(\pi) - E\left[\epsilon_t(\psi_{0,n})d_{\psi,t}(\pi)\right]\right\}$$

satisfies

$$\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) : \pi\in\Pi\} \Rightarrow^* \{\mathcal{G}_\psi(\pi) : \pi\in\Pi\}$$

a zero mean Gaussian process with *almost surely* uniformly continuous, and bounded, sample paths. Therefore, given $\mathcal{H}_\psi(\pi) \equiv E[y_{t-1}^2\exp\{-2\pi y_{t-1}^2\}] > 0\ \forall\pi\in\Pi$, $-\xi_\psi(\pi,b)$ is a non-central chi-squared process with continuous and bounded sample path. By application of Lemma 8.5 in Andrews and Cheng (2012b), $\pi^*(b)$ exists. By compactness of Π and continuity of the sample paths $\{\xi_\psi(\pi,b) : \pi\in\Pi\}$, $\pi^*(b)$ has a continuous distribution.

Finally, the convolution $\mathfrak{Z}_\psi(\pi^*(b),\lambda)$ is generally difficult to characterize, even under our simple ESTAR model, due to the complex relationship between $\mathfrak{Z}_\psi(\pi,\lambda)$ and $\xi_\psi(\pi,b)$. However, under the stated model, all other components of $\mathfrak{T}_\psi(\pi^*(b),\lambda,b)$ in $\mathcal{W}_\psi(\pi^*(b),\lambda)$ will carry over distribution continuity from $\pi^*(b)$. Thus, under the necessary assumption that $\mathfrak{Z}_\psi(\pi^*(b),\lambda)$ has a continuous distribution function *a.e.* on $\mathbb{R}$, then $\mathfrak{T}_\psi(\pi^*(b),\lambda,b)$ has a continuous distribution function *a.e.* on $\mathbb{R}$.

F.3.2 Denominator $\bar{v}^2(\pi,\lambda,b)$

Be the same arguments, $\{\tau_\beta(\pi,b) : \pi\in\Pi\}$ is a Gaussian process with *almost surely* uniformly continuous, and bounded, sample paths. Therefore $v^2(\omega^*(\pi,b),\pi,\lambda)$ has a continuous distribution *a.e.* on $\mathbb{R}$. By assumption $v^2(\omega,\pi,\lambda) > 0$ uniformly in (ω,π) for each $\lambda\in\Lambda$. Therefore $\bar{v}^2(\pi,\lambda,b) \equiv v^2(\omega^*(\pi,b),\pi,\lambda) > 0$ *a.s.* uniformly in (b,π) for each $\lambda\in\Lambda$.

F.3.3 $\mathcal{T}_\psi(\lambda,h)$

Thus, if $\mathfrak{Z}_\psi(\pi^*(b),\lambda)$ has a continuous distribution function *a.e.* on $\mathbb{R}$, then $\mathcal{T}_\psi(\lambda,b)$ has a continuous distribution *a.e.* on $\mathbb{R}$, for each b and λ. The same argument applies to the complete set of nuisance parameters h containing b.

G References

Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

Andrews, D. W. K., and X. Cheng (2012a): "Estimation and Inference with Weak, Semi-Strong and Strong Identification," *Econometrica*, 80, 2153–2211.

——— (2012b): "Supplement to 'Estimation and Inference with Weak, Semi-Strong, and Strong Indentification'," Dept. of Economics, Yale University.

——— (2013a): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics*, 173, 36–56.

——— (2013b): "Supplemental Appendix to 'Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure'," Dept. of Economics, Yale University.

Arcones, M. A., and B. Yu (1994): "Central Limit Theorems for Empirical U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, 7, 47–71.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

Dudley, R. M. (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899–929.

Galambos, J. (1987): *The Asymptotic Theory of Extreme Order Statistics*. Krieger: Malabar.

Gine, E., and J. Zinn (1990): "Bootstrapping General Empirical Measures," *Annals of Statistics*, 18, 851–869.

Hill, J. B. (2008): "Consistent and Non-Degenerate Model Specification Tests Against Smooth Transition and Neural Network Alternatives," *Annales D' Economie et de Statistique*, 90, 145–179.

——— (2013): "Consistent GMM Residuals-Based Tests of Functional Form," *Econometric Reviews*, 32, 361–383.

Hoffman-Jϕrgensen, J. (1984): "Convergence of Stochastic Processes on Polish Spaces," mimeo.

——— (1991): "Convergence of Stochastic Processes on Polish Spaces," in *Various Publications Series 39*. Aarhus Universitet, Aarhus, Denmark, mimeo.

Ibragimov, I. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

McLeish, D. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 829–839.

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer, New York.

——— (1990): "Empirical Processes: Theory and Applications," *Regional Conference Series in Probability and Statistics*, 2.

Stinchcombe, M., and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295–325.

Terasvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208 218.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge Univ. Press.

Vapnik, V. K., and A. Y. Červonenkis (1971): "On the Uniform Convergence of Relative Frequencies of Events to their Probabilities," *Theory of Probability and its Applications*, 16, 264–280.

Table 1: STAR Test Rejection Frequencies: Sample Size $n = 100$, $\sigma = 1$

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
	Strong Identification: $\beta_n = .3$								
supremum	.025	.094	.163	.147	.280	.365	.757	.872	.907
average	.025	.078	.135	.087	.209	.289	.552	.726	.804
random	.011	.052	.096	.053	.143	.232	.446	.635	.732
random LF	**.007**	**.015**	**.038**	**.013**	**.066**	**.141**	**.442**	**.553**	**.661**
random ICS-1	**.013**	**.050**	**.089**	**.028**	**.089**	**.170**	**.379**	**.593**	**.692**
PVOT^e	.015	.065	.124	.101	.257	.335	.727	.859	.883
PVOT LF	**.007**	**.014**	**.052**	**.026**	**.121**	**.208**	**.552**	**.781**	**.817**
PVOT ICS-1	**.007**	**.043**	**.073**	**.042**	**.153**	**.237**	**.622**	**.815**	**.842**
	Weak Identification: $\beta_n = .3/\sqrt{n}$								
supremum	.064	.155	.239	.337	.574	.681	.929	.978	.993
average	.057	.146	.219	.215	.430	.554	.739	.888	.932
random	.027	.083	.175	.164	.343	.474	.604	.810	.870
random LF	**.012**	**.042**	**.093**	**.060**	**.161**	**.308**	**.467**	**.685**	**.794**
random ICS-1	**.012**	**.046**	**.104**	**.116**	**.261**	**.382**	**.545**	**.749**	**.841**
PVOT	.038	.127	.196	.328	.542	.591	.893	.968	.950
PVOT LF	**.015**	**.049**	**.108**	**.108**	**.320**	**.398**	**.710**	**.911**	**.916**
PVOT ICS-1	**.014**	**.049**	**.107**	**.221**	**.435**	**.486**	**.830**	**.942**	**.932**
	Non-Identification: $\beta_n = \beta_0 = 0$								
supremum	.066	.164	.249	.358	.584	.696	.902	.970	.983
average	.062	.148	.226	.233	.438	.548	.716	.872	.911
random	.044	.107	.186	.184	.380	.505	.634	.793	.864
random LF	**.013**	**.046**	**.115**	**.069**	**.191**	**.327**	**.498**	**.725**	**.818**
random ICS-1	**.013**	**.047**	**.116**	**.137**	**.298**	**.481**	**.583**	**.769**	**.847**
PVOT	.049	.134	.190	.322	.554	.624	.890	.962	.957
PVOT LF	**.015**	**.061**	**.117**	**.122**	**.322**	**.415**	**.740**	**.911**	**.936**
PVOT ICS-1	**.015**	**.057**	**.116**	**.253**	**.464**	**.570**	**.847**	**.939**	**.954**

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *supremum* and *average* tests are based on a wild bootstrapped p-value. *random*: $\mathcal{T}_n(\lambda)$ with randomly chosen λ on [1,5]. *PVOT*: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the *least favorable* p-value, and ICS-1 is the type 1 *identification category selection* p-value with threshold $\kappa_n = \ln(\ln(n))$.

Table 2: STAR Test Rejection Frequencies: Sample Size $n = 250$, $\sigma = 1$

	H_0: LSTAR				H_1-weak				H_1-strong		
	1%	5%	10%		1%	5%	10%		1%	5%	10%
	Strong Identification: $\beta_n = .3$										
supremum	.018	.088	.163		.359	.468	.551		.953	.984	.990
average	.014	.077	.133		.262	.387	.468		.873	.949	.975
random	.014	.064	.126		.165	.299	.396		.793	.912	.952
random LF	**.001**	**.010**	**.025**		**.067**	**.235**	**.368**		**.688**	**.888**	**.936**
random ICS-1	**.008**	**.031**	**.077**		**.076**	**.244**	**.375**		**.762**	**.902**	**.947**
PVOT	.016	.067	.125		.328	.437	.517		.952	.983	.991
PVOT LF	**.004**	**.020**	**.041**		**.132**	**.348**	**.417**		**.938**	**.972**	**.976**
PVOT ICS-1	**.011**	**.051**	**.108**		**.147**	**.370**	**.433**		**.947**	**.978**	**.985**
	Weak Identification: $\beta_n = .3/\sqrt{n}$										
supremum	.051	.139	.224		.764	.922	.957		.992	1.00	1.00
average	.046	.118	.215		.539	.779	.853		.969	.992	.998
random	.027	.086	.169		.451	.695	.785		.911	.979	.993
random LF	**.018**	**.060**	**.097**		**.180**	**.481**	**.641**		**.851**	**.961**	**.980**
random ICS-1	**.018**	**.058**	**.098**		**.298**	**.633**	**.770**		**.926**	**.975**	**.991**
PVOT	.051	.122	.201		.740	.894	.934		1.00	1.00	1.00
PVOT LF	**.014**	**.061**	**.110**		**.380**	**.708**	**.805**		**.990**	**1.00**	**1.00**
PVOT ICS-1	**.015**	**.060**	**.111**		**.618**	**.848**	**.878**		**.999**	**1.00**	**1.00**
	Non-Identification: $\beta_n = \beta_0 = 0$										
supremum	.061	.152	.223		.751	.922	.956		1.00	1.00	1.00
average	.054	.145	.200		.526	.765	.849		.975	.996	.999
random	.036	.123	.184		.417	.696	.803		.025	.976	.988
random LF	**.008**	**.047**	**.108**		**.205**	**.504**	**.655**		**.838**	**.955**	**.973**
random ICS-1	**.008**	**.049**	**.109**		**.411**	**.653**	**.770**		**.923**	**.977**	**.989**
PVOT	.036	.145	.211		.732	.885	.930		1.00	1.00	1.00
PVOT LF	**.010**	**.058**	**.114**		**.373**	**.717**	**.806**		**.990**	**1.00**	**1.00**
PVOT ICS-1	**.010**	**.059**	**.116**		**.682**	**.853**	**.898**		**1.00**	**1.00**	**1.00**

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *supremum* and *average* tests are based on a wild bootstrapped p-value. *random*: $\mathcal{T}_n(\lambda)$ with randomly chosen λ on [1,5]. *PVOT*: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the *least favorable* p-value, and ICS-1 is the type 1 *identification category selection* p-value with threshold $\kappa_n = \ln(\ln(n))$.

Table 3: STAR Test Rejection Frequencies: Sample Size $n = 500$, $\sigma = 1$

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
	Strong Identification: $\beta_n = .3$								
supremum	.029	.069	.153	.441	.590	.676	.997	.999	.999
average	.022	.055	.120	.382	.546	.624	.988	.996	.997
random	.008	.049	.098	.328	.488	.598	.976	.999	.996
random LF	**.001**	**.018**	**.042**	**.227**	**.450**	**.565**	**.967**	**.989**	**.998**
random ICS-1	**.009**	**.046**	**.096**	**.230**	**.449**	**.565**	**.974**	**.990**	**.998**
PVOT	.014	.055	.115	.423	.568	.655	.996	.999	.999
PVOT LF	**.002**	**.023**	**.051**	**.311**	**.509**	**.618**	**.995**	**.998**	**1.00**
PVOT ICS-1	**.013**	**.058**	**.106**	**.314**	**.510**	**.618**	**.995**	**.998**	**1.00**
	Weak Identification: $\beta_n = .3/\sqrt{n}$								
supremum	.044	.134	.184	.984	.998	1.00	1.00	1.00	1.00
average	.029	.125	.176	.883	.968	/989	1.00	1.00	1.00
random	.032	.096	.162	.817	.929	.970	.995	.998	.998
random LF	**.009**	**.051**	**.108**	**.519**	**.835**	**.914**	**.984**	**.996**	**.998**
random ICS-1	**.009**	**.051**	**.120**	**.785**	**.921**	**.954**	**.990**	**.998**	**1.00**
PVOT	.050	.118	.194	.981	.995	1.00	1.00	1.00	1.00
PVOT LF	**.012**	**.053**	**.109**	**.823**	**.965**	**.975**	**1.00**	**1.00**	**1.00**
PVOT ICS-1	**.012**	**.054**	**.109**	**.958**	**.987**	**.993**	**1.00**	**1.00**	**1.00**
	Non-Identification: $\beta_n = \beta_0 = 0$								
supremum	.051	.151	.196	.981	.998	.998	1.00	1.00	1.00
average	.043	.136	.189	.886	.968	.984	1.00	1.00	1.00
random	.047	.111	.177	.826	.938	.967	.997	1.00	1.00
random LF	**.006**	**.058**	**.110**	**.549**	**.859**	**.926**	**1.00**	**1.00**	**1.00**
random ICS-1	**.006**	**.058**	**.109**	**.827**	**.940**	**.973**	**1.00**	**1.00**	**1.00**
PVOT	.061	.148	.208	.977	.993	.996	1.00	1.00	1.00
PVOT LF	**.014**	**.058**	**.108**	**.853**	**.970**	**.989**	**1.00**	**1.00**	**1.00**
PVOT ICS-1	**.013**	**.057**	**.107**	**.978**	**.996**	**.998**	**1.00**	**1.00**	**1.00**

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *supremum* and *average* tests are based on a wild bootstrapped p-value. *random*: $\mathcal{T}_n(\lambda)$ with randomly chosen λ on [1,5]. *PVOT*: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the *least favorable* p-value, and ICS-1 is the type 1 *identification category selection* p-value with threshold $\kappa_n = \ln(\ln(n))$.